SUPPLEMENTAL MATERIALS

METHODS

Joint models for longitudinal chimersim and event times

Denote Y_{ijs}= jth measurement of the percentage of donor cells, of type $s = T,M$ for T-cell or myeloid cells, in the peripheral blood of subject i at time t_{ij} post allo-SCT, for $i = 1, \cdots, n$ and $j = 1, \cdots n_i$. Let $U_{ijs} = I(Y_{ijs} = 100)$ be the corresponding indicator of complete (100%) donor chimerism. Suppressing the indices j, M, T, the repeated measurement process is stopped at the censoring time $C_i = \min(C_i^*, T_i)$, the minimum of the independent censoring time C_i^*, and the dependent terminal event (death or relapse) time T_i. Denote by Z_i the time-independent baseline covariate vector for subject i. The first joint model links the longitudinal binary indicator of complete donor chimerism with event time through a shared random patient effect, α_i^d, assumed to be distributed iid N(0, σ^2) for all patients. The joint model is specified by the longitudinal probability of complete donor T-cell chimerism, π_{ijT} = $\Pr(U_{ijT}=1 \mid Z_i, \alpha_i^d)$, and a Cox model, as follows:

Longitudinal mixed model : $\text{logit}(\pi_{ijT}) = \gamma_0 + \gamma_1 t_{ij} + Z_i \gamma_2 + \alpha_i^d$.

Cox model : $\lambda_i(t) = \lambda_0(t) \exp(Z_i^T \beta + \delta_1 \alpha_i^d)$

where $\lambda_0(t)$ is the baseline hazard function. The parameter δ_1 is the effect of the random patient effect on the hazard of death. Only the indicators U_{ijT} of complete T-cell chimerism

were used because these were nearly identical to the indicators U_{ijM} of complete myeloid cell chimerism.

Joint model for longitudinal percentages of donor chimerism and event time

Longitudinal plots of percent T-cell or myeloid cell donor chimerism (Figure 2) suggest that subjects whose percentage of donor cells decreased rapidly tended to die earlier. We therefore fit the following joint shared random effect model. Let $I(s = M)$ be the indicator of myeloid chimerism, μ_i and β_i normally distributed subject-specific random intercept and slope for change in percent chimerism over time, with ε_{ij} a residual error term assumed normally distributed with mean zero and variance σ_ε^2 . The joint model is specified as follows :

Longitudinal mixed model : $Y_{ijs} = \mu_i + \beta_i t_{ij} + \beta_T I(s = M) + Z_i^T \beta_Z + \varepsilon_{ij}, s = T, M$

Cox model : $\lambda_i(t) = \lambda_0(t) \exp\{Z_i^T \gamma + \delta_1 \beta_i^+ + \delta_2 \beta_i^-\}$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the unspecified baseline hazard function, $\beta_i^+ = \beta_i I(\beta_i \geq 0)$ and $\beta_i^- = -\beta_i I(\beta_i < 0)$. In this longitudinal model, a positive (negative) value of β_i means that the chimerism increased (decreased) over time. The parameter β_i^+ equals β_i if β_i is positive, and otherwise equals zero. The parameter of β_i^- equals the absolute value of β_i if β_i is negative, and otherwise equals zero. In this model, decreasing

and increasing paths of percent donor cells over time may have different effects on survival, represented by the two parameters δ_1 and δ_2 , which quantify the respective effects of β_i^+ and β_i^- on the risk of relapse/death.

Supplemental Table 1: Fitted joint model for longitudinal complete T-cell chimerism indicator and DFS time.

	Coefficient	SE	P-value	95% CI
Longitudinal model for complete donor T-cell chimerism[1]				
Time after allo-SCT	0.089	0.019	<0.001	(0.052, 0.126)
Disease status at time of allo-SCT				
CR vs. Active	0.469	0.477	0.327	(-0.466, 1.404)
Allotype				
MRD vs. MUD	0.016	0.792	0.984	(-1.536, 1.568)
Mismatched vs. MUD	-0.718	0.492	0.146	(-1.094, 0.736)
Cytogenetic risk group				
Good vs. poor	-0.179	0.467	0.702	(-1.094, 0.736)
Intermediate vs. poor	0.070	0.827	0.933	(-1.551, 1.691)
Cox model for DFS time[2]				
Shared random effect	**-0.244**	**0.075**	**0.001**	**(-0.391, -0.097)**
Disease status at time of allo-SCT				
CR vs. Active	-1.055	0.233	<0.001	(-1.511, -0.598)
Allotype				
MRD vs. MUD	-0.329	0.244	0.180	(-0.807, 0.149)
Mismatched vs. MUD	0.243	0.333	0.467	(-0.410,0.900)
Cytogenetic risk group				
Good vs. poor	-0.394	0.368	0.286	(-1.115, 0.327)
Intermediate vs. poor	-0.415	0.231	0.073	(-0.868, 0.038)

[1]A positive (negative) coefficient estimate in the longitudinal model corresponds to higher (lower) probability of complete T-cell chimerism.

[2]A positive (negative) coefficient estimate in the Cox model corresponds to a higher (lower) risk of disease progression and thus on average a shorter (longer) DFS time.

Supplemental Table 2: Fitted joint model for longitudinal chimerism percentage and DFS time.

	Coefficient	SE	P-value	95% CI
Longitudinal model for T-cell or myeloid cell percent donor cells[1]				
Type of chimerism				
Myeloid vs. T-cell	9.227	0.840	<0.001	(7.581,10.873)
Time after allo-SCT	0.182	0.346	0.600	(-0.496, 0.860)
Disease status at time of allo-SCT				
CR vs. Active	0.941	1.496	0.530	(-1.991, 3.873)
Allotype				
MRD vs. MUD	-5.618	1.583	0.001	(-8.721, -2.515)
Mismatched vs. MUD	-3.602	2.511	0.153	(-8.523, 1.320)
Cytogenetic risk group				
Good vs. poor	0.262	1.596	0.869	(-2.866, 3.390)
Intermediate vs. poor	0.798	2.678	0.766	(-4.451, 6.047)
Cox model for DFS time[2]				
Positive Slope (Increasing donor chimerism)	**0.181**	**0.206**	**0.379**	**(-0.223, 0.585)**
Negative Slope (Decreasing donor chimerism)	**0.163**	**0.069**	**0.019**	**(0.028, 0.298)**
Disease status at time of allo-SCT				
CR vs. Active	-0.709	0.206	0.001	(-1.112, -0.305)
Allotype				
MRD vs. MUD	-0.080	0.243	0.742	(-0.556, 0.396)
Mismatched vs. MUD	0.223	0.323	0.490	(-0.410, 0.856)
Cytogenetic risk group				
Good vs. poor	-0.277	0.216	0.199	(-0.700, 0.146)
Intermediate vs. poor	-0.779	0.410	0.058	(-1.583,0.025)

[1]A positive (negative) coefficient estimate in the longitudinal model corresponds to higher (lower) percent donor cells.

[2]A positive (negative) coefficient estimate in the Cox model corresponds to a higher (lower) risk of disease progression and thus on average a shorter (longer) DFS time.