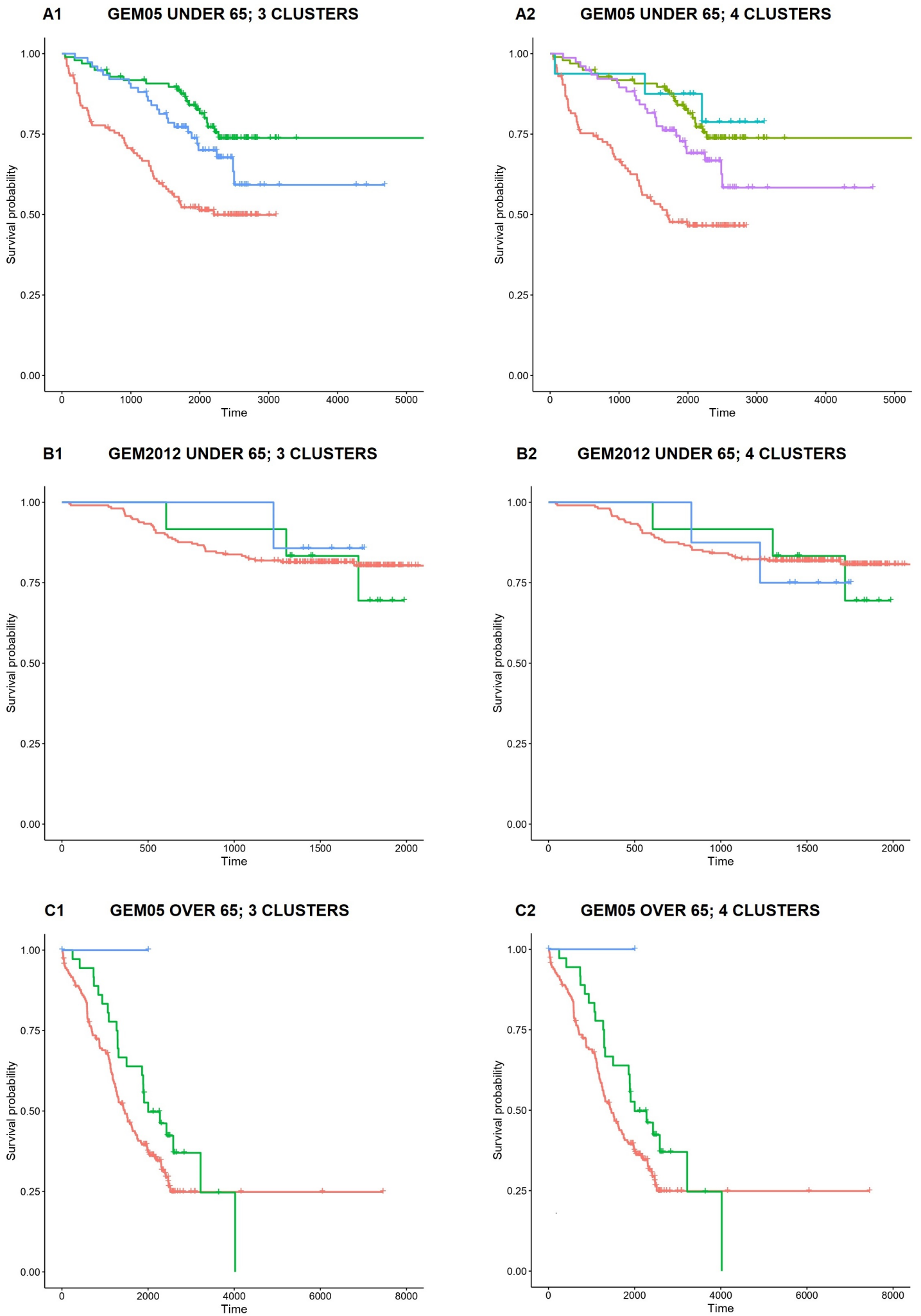
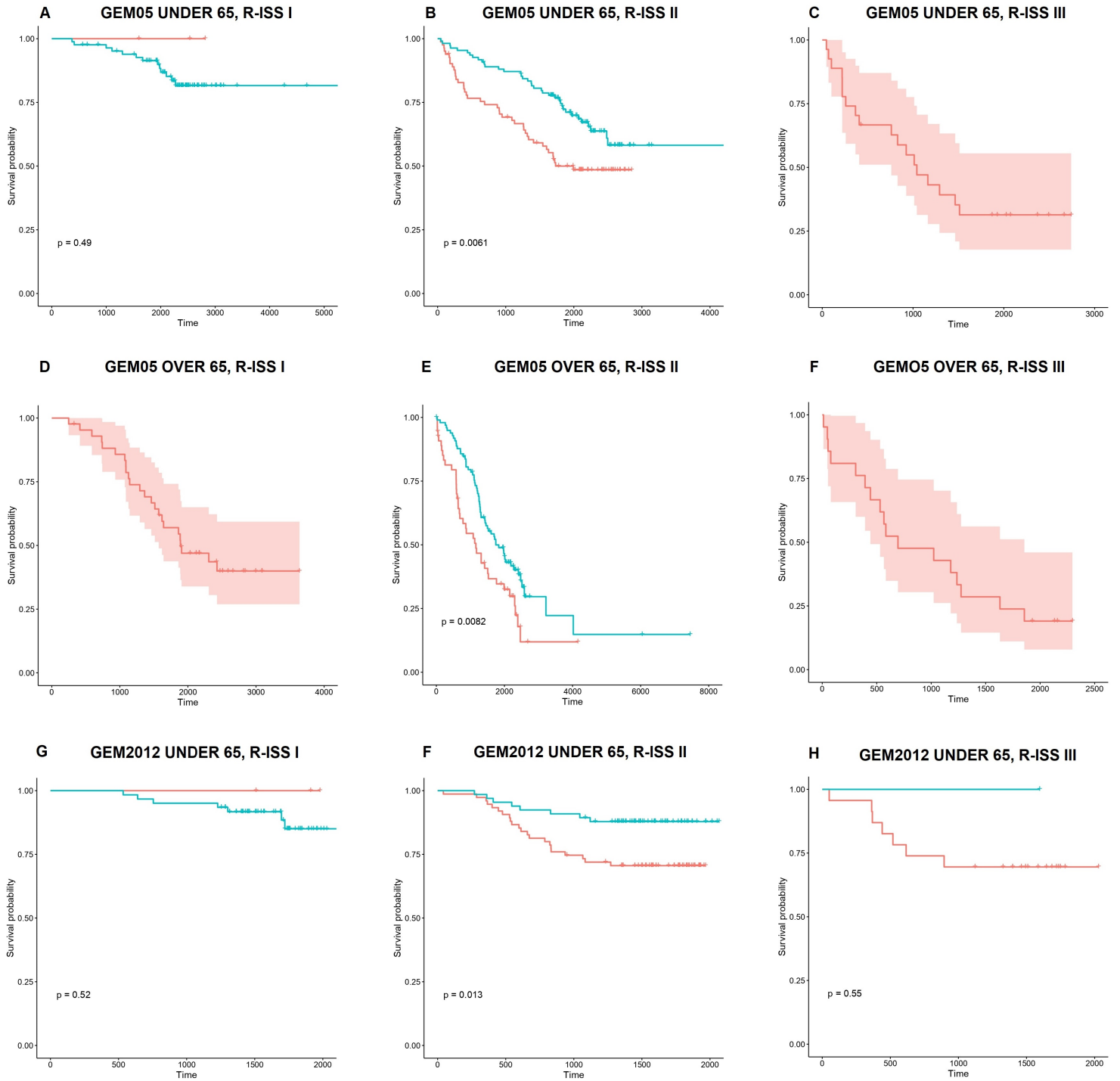


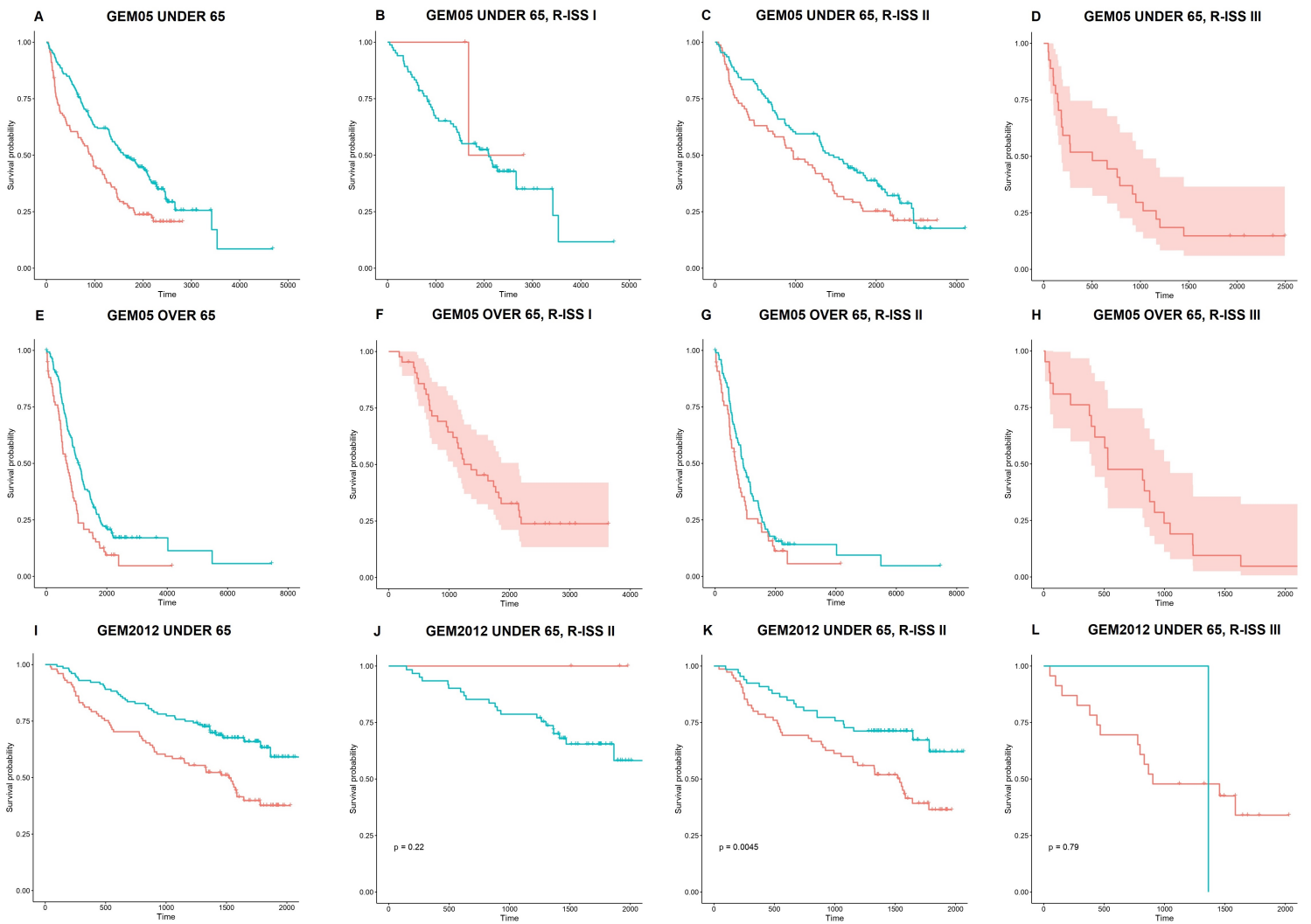
Supplementary Figure 1. Results of GMM-EM-based clusterization with 3 and 4 optimal clusters across the 3 trial cohorts.



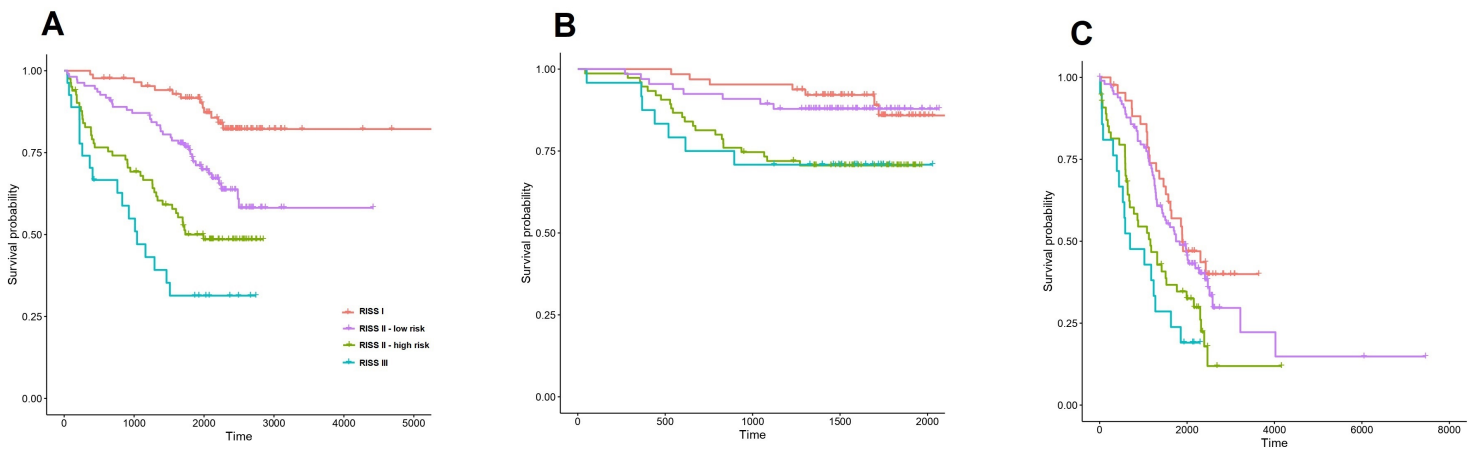
Supplementary Figure 2. Kaplan-Meier plots representing the overall survival of R-ISS I, II and III subgroups in the GEM05 under 65 years (A-C), GEM05 over 65 years (D-F) and GEM2012 under 65 years (G-H) trials. The red line represents patients in the “high-risk” cluster (cluster I), and the blue line represents patients in the “low-risk” cluster (cluster II).



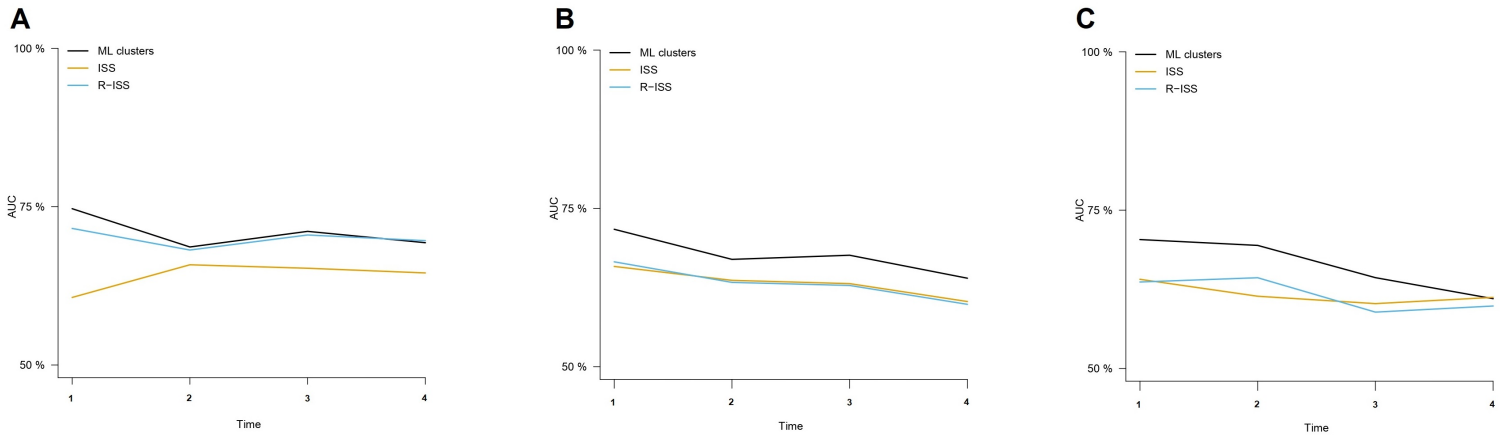
Supplementary Figure 3. Kaplan-Meier plots representing the overall survival of all patients as well as that of R-ISS I, II and III subgroups in the GEM05 under 65 years (A-D), GEM05 over 65 years (E-H) AND GEM2012 under 65 years (I-L) trials. The red line represents patients in the “high-risk” cluster (cluster I), and the blue line represents patients in the “low-risk” cluster (cluster II).



Supplementary Figure 4. Kaplan-Meier plots representing the overall survival of R-ISS I, II-low risk, II-high risk and III subgroups in the GEM05 under 65 years (A), GEM2012 under 65 years (B) and GEM05 over 65 years (C) trials.



Supplementary Figure 5. Time-dependent AUCs representing the precision of the different models to predict overall survival.



	GEM05 under 65	GEM05 over 65	GEM2012 under 65
<i>Hemoglobin*</i>	p-val 1.02 x 10 ⁻⁵ [-1.4, -0.60]	p-val 2.82 x 10 ⁻⁵ [-1.40, -0.50]	p-val 4.98 x 10 ⁻⁴ [-1.60, -0.30]
<i>Albumin-adjusted calcium*</i>	p-val 0.84 [-0.26, 0.30]	p-val 0.44 [-0.20, 0.46]	p-val 0.24 [-0.30, 0.09]
<i>B2-microglobulin*</i>	p-val 8.45 x 10 ⁻⁵ [0.45, 1.37]	p-val 5.15 x 10 ⁻⁷ [1.01, 2.32]	p-val 2.07 x 10 ⁻⁴ [0.40, 1.47]
<i>High risk cytogenetics**</i>	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵
<i>t(4;14)**</i>	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵
<i>t(14;16)**</i>	p-val < 1 x 10 ⁻⁵	p-val 0.12	p-val < 1 x 10 ⁻⁵
<i>17p del**</i>	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵
<i>Raised LDH**</i>	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵	p-val < 1 x 10 ⁻⁵
<i>Durie-Salmon stages I, II and III***</i>	Cluster 1: 3.30%, 77.05%, 50.41% Cluster 2: 7.78%, 52.85% & 39.38%	Cluster 1: 7.89%, 38.16% & 53.94% Cluster 2: 7.04%, 59.15% & 33.80%	Cluster 1: 12.87%, 33.66% & 53.47% Cluster 2: 8.59%, 42.19% & 49.22%

* Wilcoxon Rank-Sum Test
plus 95% Confidence Interval

**Fisher's Test

***Descriptive

Supplementary Table 1.

Differential distribution between the 2 unsupervised clusters of the different predictive variables which compose it. Continuous variables were analyzed with the Wilcoxon Rank-Sum Test and binomial variables were analyzed with Fisher's test. For Durie-Salmon stage distribution, descriptive statistics are provided.

Variables	GEM05 under 65		GEM05 over 65		GEM2012 under 65	
	<i>p-val</i>	<i>HR (95%CI)</i>	<i>p-val</i>	<i>HR (95%CI)</i>	<i>p-val</i>	<i>HR (95%CI)</i>
<i>II-cluster 2 vs I</i>	2.00 x 10 ⁻³	2.71 [1.44, 5.11]	0.42	1.21 [0.75, 1.95]	0.68	1.23 [0.45, 3.42]
<i>III vs II-cluster 1</i>	0.06	1.69 [0.97, 2.95]	0.21	0.96 [0.39, 1.23]	0.90	0.95 [0.40, 2.22]

Supplementary Table 2. Cox regression results comparing the outcomes of R-ISS I vs II-low risk and R-ISS III vs II-high risk.

	GEM05 under 65	GEM05 over 65	GEM2012 under 65
<i>N</i>	191	154	141
<i>Cluster I vs II</i>	42.93% vs 57.07%	35.71% vs 64.29%	41.13% vs 58.87%
<i>High Risk Cytogenetics Cluster I vs II</i>	52.43% vs 0%	47.27% vs 0%	56.63% vs 0%
<i>Durie-Salmon stages: I, II & III</i>	Cluster 1: 3.66%, 46.34%, 50.00% Cluster 2: 6.42%, 46.79%, 46.79%	Cluster 1: 10.91%, 41.82%, 47.27% Cluster 2: 6.06%, 51.52%, 42.42%	Cluster 1: 9.64%, 33.73%, 56.63% Cluster 2: 5.17%, 41.38%, 53.45%
<i>Median Hemoglobin (g/dL)</i>	10.5 vs 10.3	9.9 vs 10.2	10.4 vs 11.0
<i>Median albumin-adjusted calcium (mg/dL)</i>	9.73 vs 9.88	9.88 vs 10.18	9.5 vs 9.6
<i>Median B2-microglobulin (mg/dL)</i>	3.35 vs 4.09	4.3 vs 4.3	3.5 vs 4.36
<i>Raised LDH</i>	41.46% vs 0%	27.27% vs 0%	27.71% vs 0%
<i>Raised LDH or high risk cytogenetics</i>	86.59% vs 0%	69.09% vs 0%	72.29% vs 0%
<i>Survival of Cluster I patients with and without either high risk cytogenetics or raised LDH</i>	HR 1.12, p-value 0.81	HR 1.33, p-value 0.41	HR 0.98, p-value 0.96

Supplementary Table 3.

Distribution of the variables included in the unsupervised machine learning model according to cluster membership for R-ISS 2 patients. Also represented are the results of the survival analysis between "high-risk" cluster I patients with and without either high risk cytogenetics or raised LDH.

	GEM05 under 65	GEM2012 under 65	GEM05 over 65
<i>ISS</i>	0. 623	0. 598	0. 578
<i>RISS</i>	0. 648	0. 600	0. 563
<i>UNSUPERVISED MODEL</i>	0. 645	0. 653	0. 591
<i>ISS+RISS</i>	0. 615	0. 592	0. 577
<i>ISS + UNSUPERVISED MODEL</i>	0. 690	0. 659	0. 619
<i>RISS + UNSUPERVISED MODEL</i>	0. 688	0. 641	0. 610
<i>ISS + RISS + UNSUPERVISED MODEL</i>	0. 700	0. 644	0. 615

Supplementary Table 4.

C-indexes obtained after 10-fold cross-validation of the different cox regression models.