Supplementary material

The pre-registered study analysis plan, SAP (*cf*. reference [10])

The role of heart-related dose-volume metrics on overall survival in the RTOG 0617 clinical trial

Study Analysis Plan (SAP id: MSKSPI1)

Maria Thor
May 16th 2018

1 Study scope

1.1 Investigators

Maria Thor, Joseph O Deasy, Chen Hu, Hak Choy, Ritsuko Komaki, Gregory Masters, George R Blumeschein, Kenneth Forster, Jung Hun Oh, Vivek Kavadi, Samir Narayan, Robert Timmerman, Clifford Robinson, Joel S Greenberger, David Biggs, Mark Augspurger, Joanne Meng, and Jeffrey Bradley

1.2 Aim and inclusion criteria

This dose-response focused study will include all stage III non-small cell lung cancer (NSCLC) patients with complete dose-volume and follow-up data from the 544 patients accumulated in the four randomization arms in the RTOG 0617 trial (last patient accrued on Nov 22nd 2011)[1]. The primary outcome measure is overall survival defined from the start date of randomization to the date of death (right-censoring if alive at last follow-up visit).

2 Analysis description

2.1 Input data and parameterization

All data will be randomly split into a model training (70%) and a model validation (30%) subset. The prescription dose level distribution will remain similar between the subsets by repeats random selection. The holdout validation subset will only be used to assess performance of derived training models. Input data will be dose-volume histograms (DVHs) for the atria, pericardium, ventricles, and the tumor-subtracted lung, as well as available and relevant disease, patient and treatment characteristics. The DVHs for all substructures will be parameterized as the minimum and the mean dose to the hottest x % volume (Dx, MOHx).

2.2 Dose-response modelling

2.2.1 Methodology I

Cox proportional Hazard's analysis will be the primary method of analysis. Variables with a median Bonferroni-corrected univariate p-value$\leq$0.05 averaged over 1000 bootstrap datasets will be considered candidate predictors with one best (lowest p-value) dose variable considered per substructure. These candidate predictors will be passed on to multivariate Cox proportional Hazard's analysis (MVA), which will be

performed using forward-stepwise selection, and a candidate predictor will be considered in the final MVA model if $p<0.05$ of the log-likelihood ratio statistics.

Similar to univariate analysis, bootstrapping with 1000 datasets that are randomly selected will be considered, *i.e.*, 1000 subpopulations will be generated, and MVA will be repeated for each subpopulation. This will result in a collection of models, and the most frequently selected model(s), *i.e.*, selected ≥10% of the times, will be considered final. After defining the final model(s), we will test the significance of interaction terms between variables in the final multivariate model(s). Ultimately, the final model(s) will be externally explored against the holdout validation data, and validation procedures will be adopted from Royston and Altman[2].

2.2.2 Methodology II

In a subsequent approach and building upon the final model(s), a second "robust" ensemble approach[3] will be considered. The bootstrapping process above will be repeated, but instead of selecting one final model of the potential collection of models, these models will be 'bagged' as follows: the collection of models will be refit to all the training data. To determine the final decision based on the prognostic index[2] of the final models, a Principle Component Analysis (PCA) will then be conducted on the vector of predicted outcomes of these models. Then, the first few PCs (the total number of PCs to include will also be addressed) will be used as predictors again using a Cox proportional Hazard's analysis. The final predictor will be an ensemble (linear sum) of a few PCs with final weights fit to the training data. Similarly to the first approach, this ensemble model will finally be explored in the holdout validation data.

3 Bibliogroahy

[1] Bradley JD, *et al*. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patient with stage IIIA or IIIB non-small cell lung cancer (RTOG 0617): a rabomized, two-by-two factorial phase 2 study. *Lancet Oncol* 2015; 16:187-99

[2] Royston T, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Medical Research Methodology* 2013; 33:1-15

[3] Zhang J. Developing robust non-linear models through bootstrap aggregated neural networks. *Neurocomputing* 1999; 25:93-113

Figures

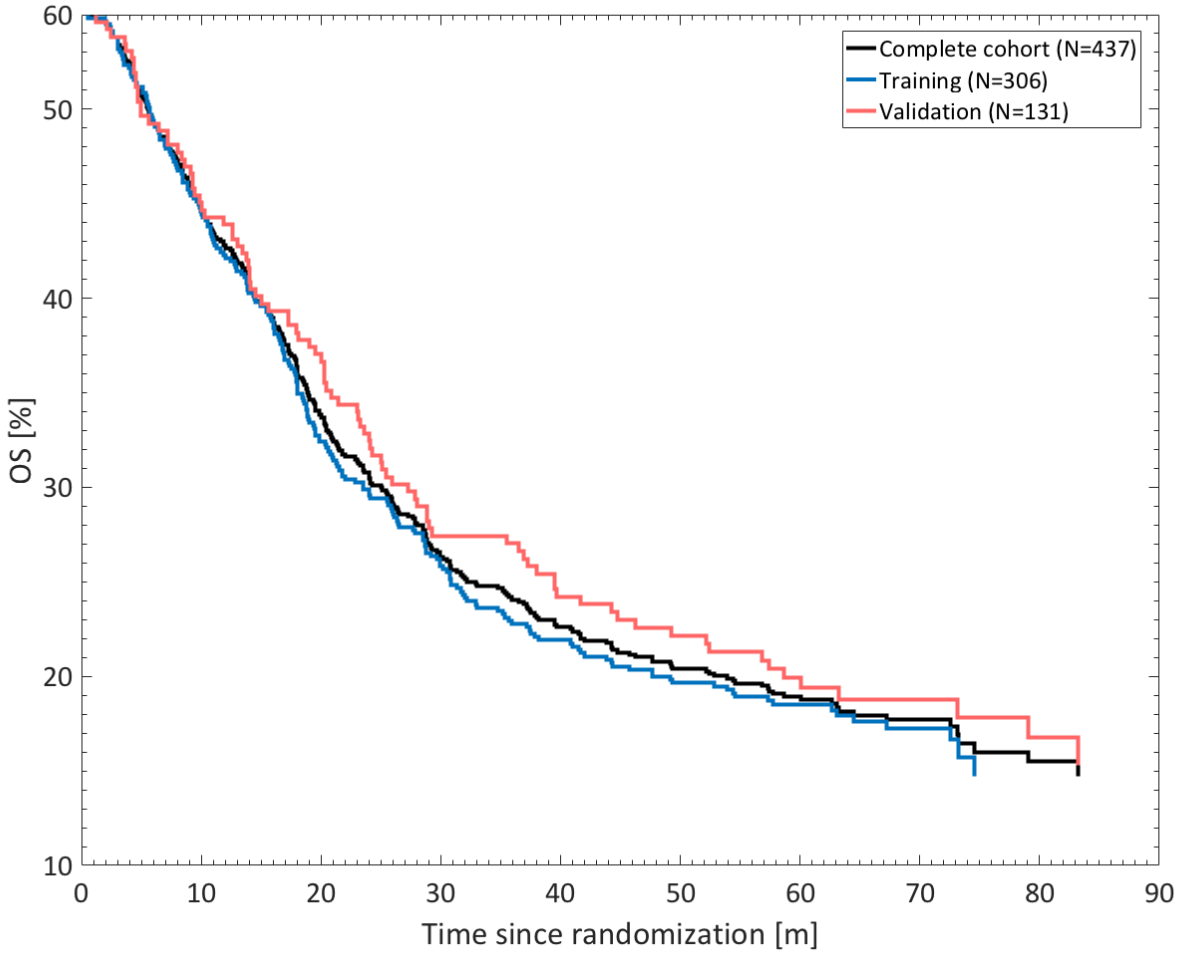

Fig S1. Overall survival curves in the complete cohort (black), in training (blue) and in validation (red). *Note: The training and validation survival curves are not significantly different (p=0.33; log-rank test; cf. Table 1).*

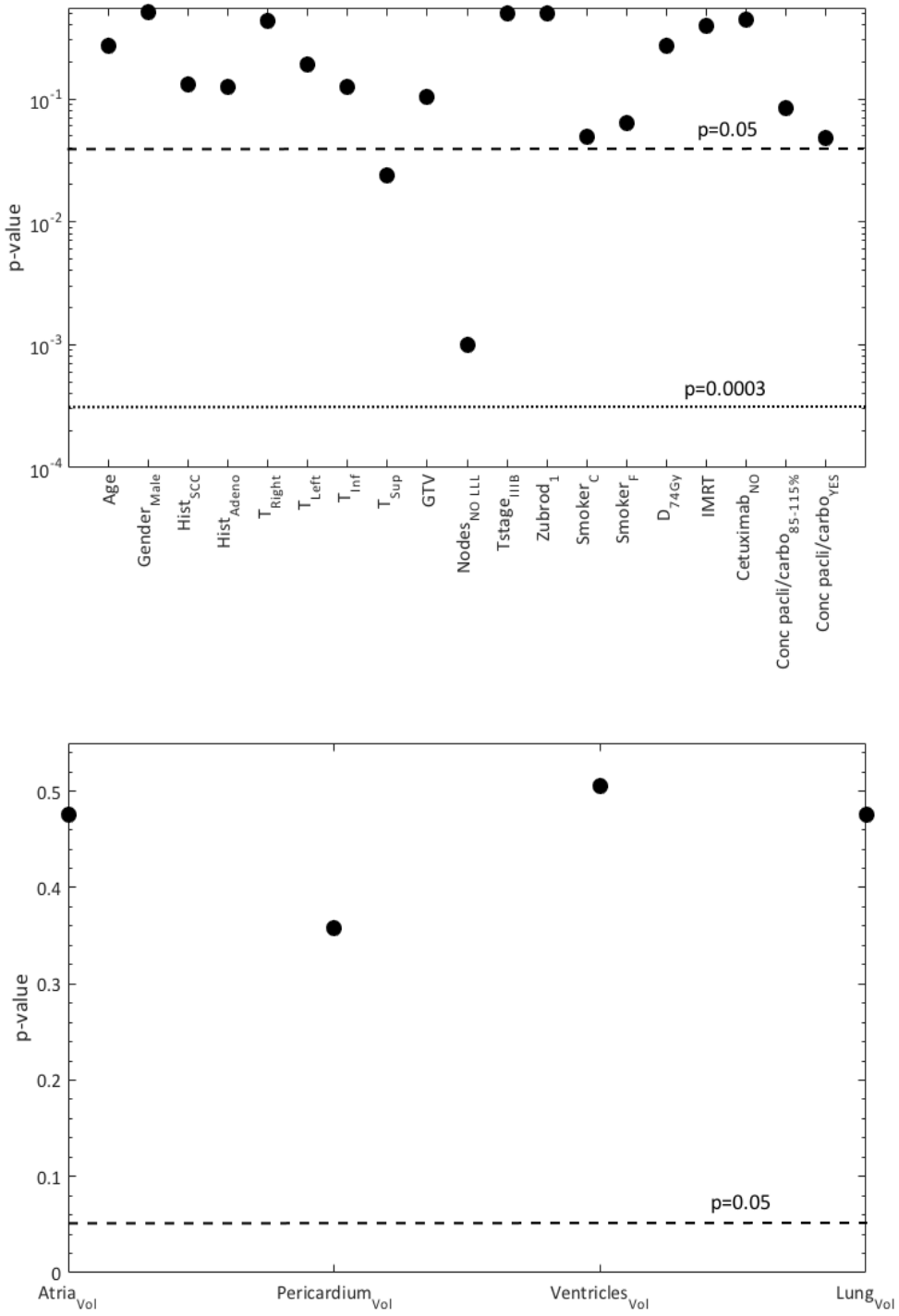

Fig S2. Univariate p-values (median over all samples) for the investigated disease, patient and treatment characteristics (upper), as well as structure volumes (lower). *Note: The dotted black lines represent the Bonferroni-corrected significance level at p=0.0003; dashed black lines are non-corrected p=0.05 levels. Abbreviations: C: Current; Carbo: Carboplatin; D: Prescribed dose; F: Former; LLL: Left lower lobe; Pacli: Paclitaxel; SCC: Squamous cell carcinoma; T: Tumor location.*

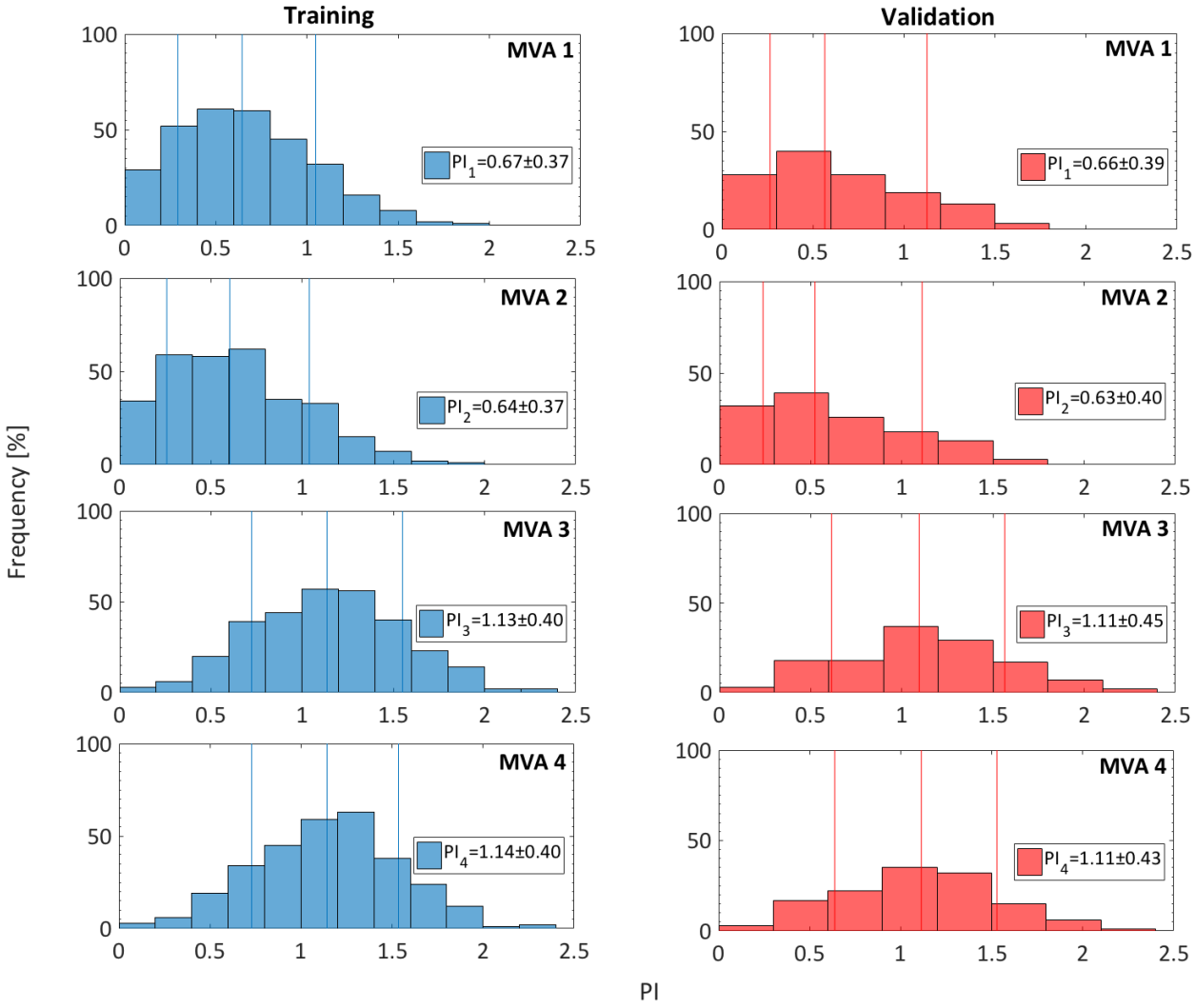

Fig S3. Prognostic index (PI) for each MVA model (upper-lower: MVA 1-4) in training and validation (left and right). *Note: The average ± standard deviation of the PI is inserted in the upper right corner for each MVA model, and the vertical blue and red lines indicate the risk group splits at the 16th, 50th and 84th centiles.*

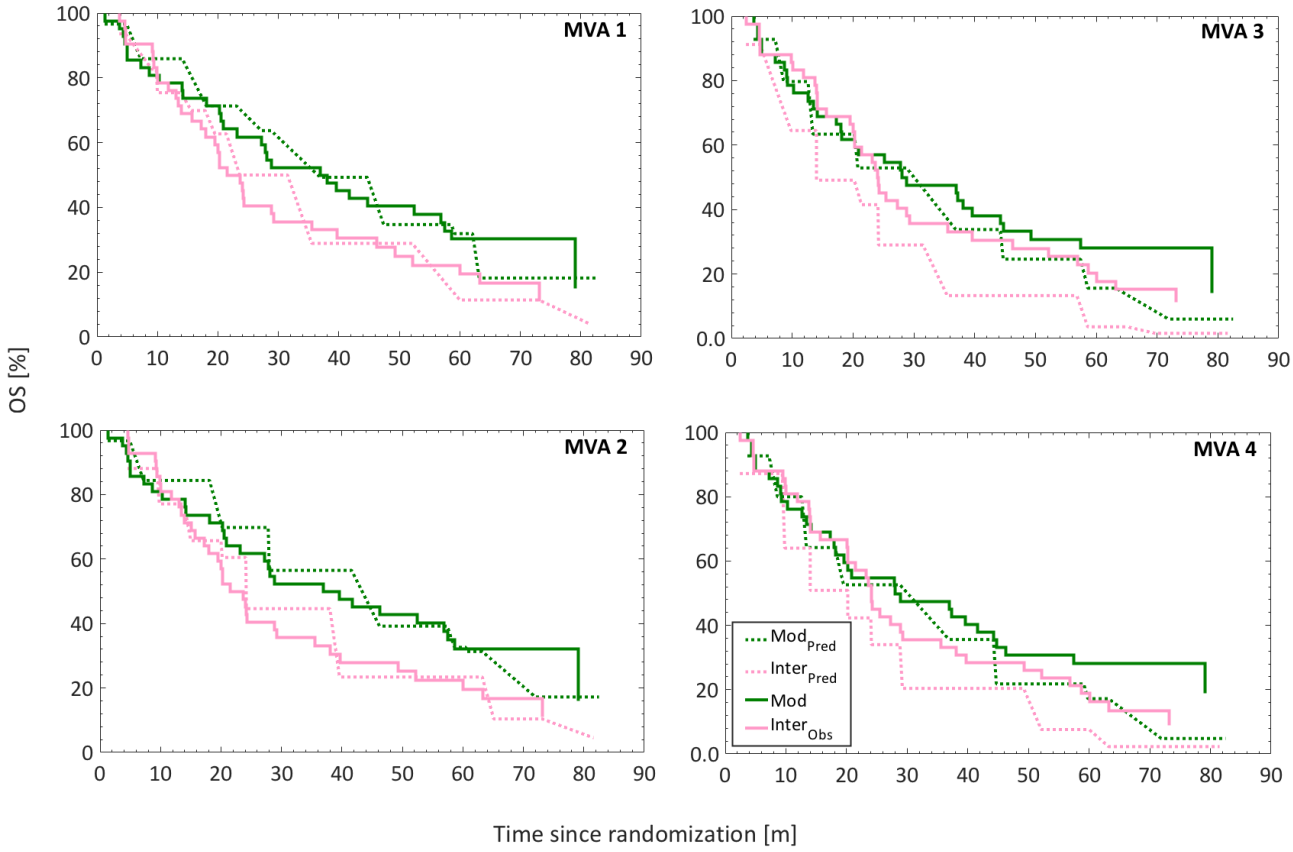

Fig S4. Kaplan-Meier curves for the moderate (Mod) and the intermediate (Inter) risk groups (*i.e.*, 16th-50th percentile and 50th-84th percentile) based on MVA models 1-4 comparing the observed survival rates (solid) vs. the predicted survival rates (dotted) in validation; the latter modifying the observed survival curve in validation based on the PI from training.

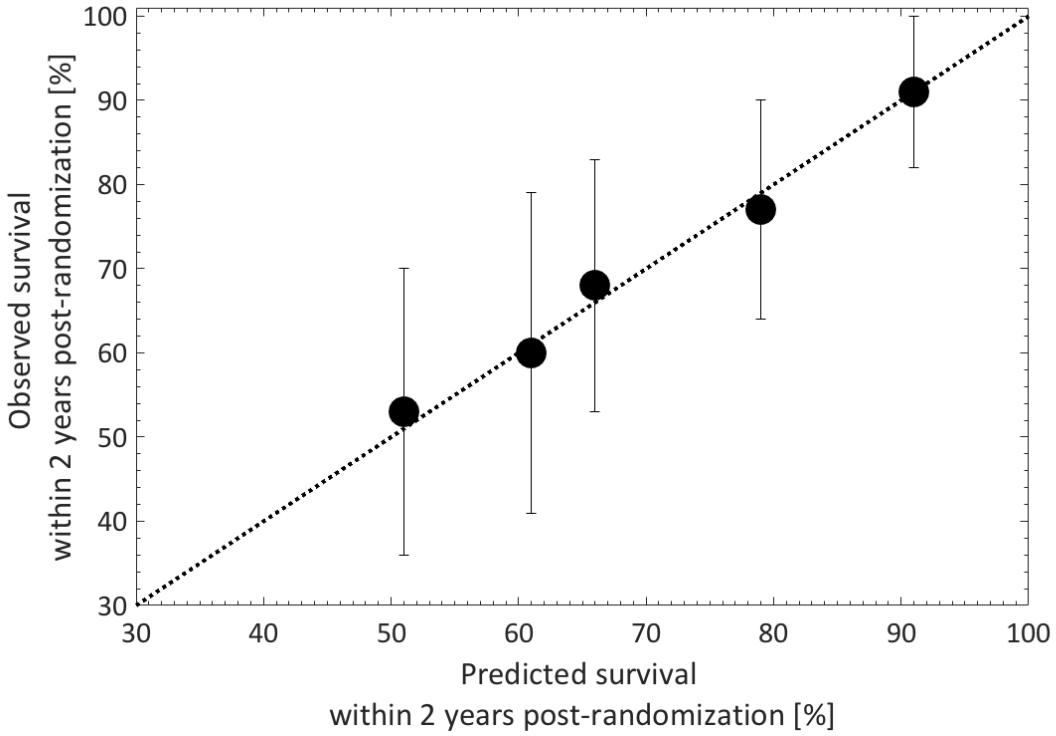

Fig S5. Calibration curve illustrating observed (y-axis) OS and by the ensemble model predicted OS (x-axis) in quintiles within two years post-randomization. *Note: The dotted line is the identity line; error bars represent standard deviations of the observed survival in each bin.*

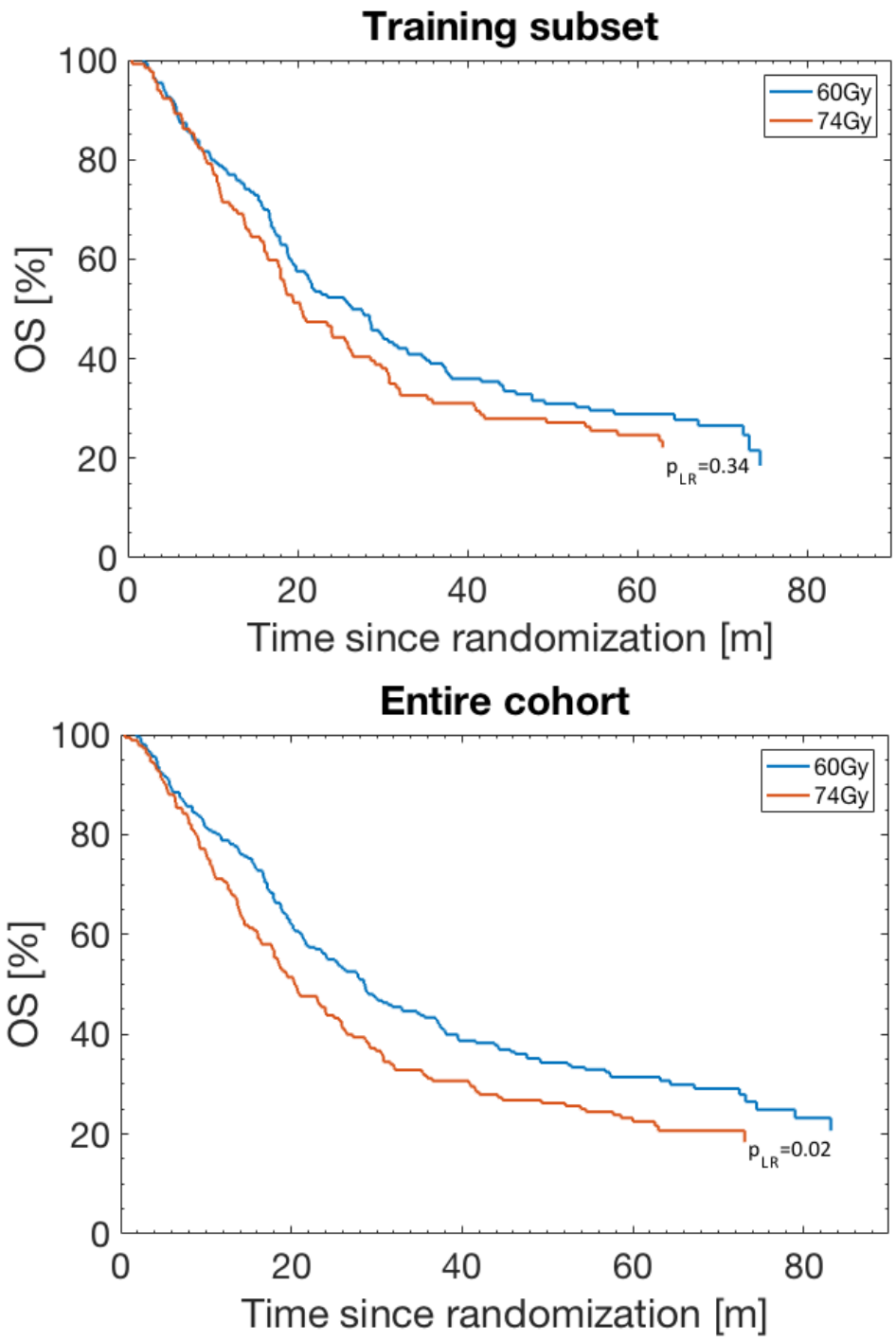

Fig S6. Kaplan-Meier curves stratified for prescription dose in the training subset (upper) and in the entire cohort (lower). *Note: The p-value comes from a log-rank test without any resampling.*

Tables

Table S1. Regression coefficients for each variable in the multivariable model (upper: univariate; lower: multivariable). These regression coefficients can be used for validation in any external cohort following the procedures outlined in Royston and Altman [13]. *Note: Univariate coefficients are given as the median value across the 1000 bootstrap samples; multivariable model coefficients have been re-fitted to the entire training data.*

Model	Univariate Analysis	Coefficient
UVA 1	Atria D45%[Gy]	0.02
UVA 2	Pericardium MOH55%[Gy]	0.03
UVA 3	Ventricles MOH5%[Gy]	0.01
UVA 4	Lung Mean[Gy]	0.09
	Multivariable Analysis	
MVA 1	Atria D45%[Gy], Pericardium MOH55%[Gy]	0.005, 0.02
MVA 2	Atria D45%[Gy], Pericardium MOH55%[Gy], Ventricles MOH5%[Gy]	0.007, 0.01, 0.004
MVA 3	Atria D45%[Gy], Pericardium MOH55%[Gy], Lung Mean[Gy]	0.006, 0.01, 0.04
MVA 4	Pericardium MOH55%[Gy], Ventricles MOH5%[Gy], Lung Mean[Gy]	0.01, 0.003, 0.04
Ensemble	Pericardium MOH55%[Gy], Atria D45%[Gy], Ventricles MOH5%[Gy], Lung Mean[Gy]	0.02, 0.002, 0.002, 0.03

Table S2. The observed (Obs.) and predicted (Pred.) survival rates for the four MVA models in validation 18 months, and 36 months after randomization.

Survival rates [%]	**MVA 1**		**MVA 2**		**MVA 3**		**MVA 4**	
18 months	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
Low risk	76	77	76	82	81	78	81	72
High risk	33	50	38	40	38	22	48	29
Moderate risk	75	71	73	84	73	63	68	64
Intermediate risk	64	70	67	64	66	49	67	51
36 months	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
Low risk	56	64	57	64	56	55	56	64
High risk	26	32	19	33	24	9	29	51
Moderate risk	54	51	51	57	53	36	52	53
Intermediate risk	38	29	40	45	40	13	38	20