Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

Modeling the Impact of Cardiopulmonary Irradiation on Overall Survival in NRG Oncology Trial RTOG 0617

Maria Thor[1], Joseph O. Deasy[1], Chen Hu[2], Elizabeth Gore[3], Voichita Bar-Ad[4], Clifford Robinson[5], Matthew Wheatley[6], Jung Hun Oh[1], Jeffrey Bogart[7], Yolanda I. Garces[8], Vivek S. Kavadi[9], Samir Narayan[10], Puneeth Iyengar[11], Jacob S. Witt[12], James W. Welsh[13], Cristopher D. Koprowski[14], James M. Larner[15], Ying Xiao[16], and Jeffrey Bradley[17]

ABSTRACT

Purpose: To quantitatively predict the impact of cardiopulmonary dose on overall survival (OS) after radiotherapy for locally advanced non–small cell lung cancer.

Experimental Design: We used the NRG Oncology/RTOG 0617 dataset. The model building procedure was preregistered on a public website. Patients were split between a training and a set-aside validation subset ($N = 306/131$). The 191 candidate variables covered disease, patient, treatment, and dose-volume characteristics from multiple cardiopulmonary substructures (atria, lung, pericardium, and ventricles), including the minimum dose to the hottest x% volume (Dx%[Gy]), mean dose of the hottest x% (MOHx%[Gy]), and minimum, mean (Mean[Gy]), and maximum dose. The model building was based on Cox regression and given 191 candidate variables; a Bonferroni-corrected P value threshold of 0.0003 was used to identify predictors. To reduce overreliance on the most highly correlated variables, stepwise multivariable analysis (MVA) was repeated on 1000 bootstrapped replicates. Multivariate sets selected in ≥10% of replicates were fit to the training subset and then averaged to generate a final model. In the validation subset, discrimination was assessed using Harrell c-index, and calibration was tested using risk group stratification.

Results: Four MVA models were identified on bootstrap. The averaged model included atria D45%[Gy], lung Mean[Gy], pericardium MOH55%[Gy], and ventricles MOH5%[Gy]. This model had excellent performance predicting OS in the validation subset ($c = 0.89$).

Conclusions: The risk of death due to cardiopulmonary irradiation was accurately modeled, as demonstrated by predictions on the validation subset, and provides guidance on the delivery of safe thoracic radiotherapy.

Introduction

Although radiotherapy is an established treatment modality for locally advanced non–small cell lung cancer (LA-NSCLC), progress has been slow, with patients being deceased on average between 1–2 years after completion of radiotherapy (1). Long-term results have recently improved fusing radiotherapy with immunotherapy checkpoint inhibitors (2), but the exact details of optimal radiotherapy remain unclear. Thus, interest continues in understanding and optimizing dose and fractionation parameters in this patient cohort.

In RTOG's 0617 randomized phase III dose escalation trial for this patient population, Bradley and colleagues (1) found, unexpectedly, that patients in the accelerated 74 Gy arm were at a higher risk of death than the patients in the conventional 60 Gy arm. In their initial analysis using heart dose-volume characteristics, they found that patients with larger heart volumes irradiated to at least 5 Gy (V5Gy[%]) were at higher risk of death.

Since the publication of the multivariate overall survival (OS) model from the RTOG 0617 trial, at least six cohort studies have presented varying heart and/or lung dose OS models after radiotherapy for LA-NSCLC (3–8). Tucker and colleagues (3) found an association with the mean lung dose, Mean[Gy], but references (4–7) instead established associations with heart dose only: heart V2Gy[%] in (4), the base of the heart in (5), left atrial V63-69Gy[%] in (6), and heart V50Gy[%] in (7). In contrast, Speirs and colleagues (8) established a simultaneous heart and lung dose association including heart V50Gy[%], heart volume, and lung V5[%]. While the majority of the datasets in (3–8) included a large number of patients (78–1,101 patients), these were typically treated at single institutions in which the variability in dose-volume variables is limited given uniform treatment procedures and techniques (9).

The aim of this study was to generate a predictive model of patient-specific risk of death based on the multi-institutional RTOG 0617 data, addressing a broad range of input data describing dose to a wide range components of the cardiopulmonary system, as well as tumor and individual characteristics. This dataset is valuable for predictive modeling, given the randomization of prescription dose, as well as multi-institutional participation, which reduces correlations due to single institutional treatment approaches (9).

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York. [2]NRG Oncology Statistics and Data Management Center, Philadelphia, Pennsylvania. [3]Zablocki Veterans Administration Medical Center, Milwaukee, Wisconsin. [4]Thomas Jefferson University Hospital, Philadelphia, Pennsylvania. [5]Siteman Cancer Center at Washington University, St. Louis, Missouri. [6]Mercy San Juan Medical Center Dignity Health, Carmichael, California. [7]State University of New York Upstate Medical University, Syracuse, New York. [8]Mayo Clinic, Rochester, Minnesota. [9]Texas Oncology Cancer Center Sugar Land, Sugar Land, Texas. [10]Saint Joseph Mercy Hospital, Ypsilanti, Michigan. [11]UT Southwestern/Simmons Cancer Center, Dallas, Texas. [12]University of Wisconsin-Madison (accruals under Washington University), Madison, Wisconsin. [13]University of Texas MD Anderson Cancer Center, Houston, Texas. [14]Christiana Care Health System-Christiana Hospital, Newark, Delaware. [15]University of Virginia Cancer Center, Charlottesville, Virginia. [16]University of Pennsylvania, Philadelphia, Pennsylvania. [17]Emory University School of Medicine, Atlanta, Georgia.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Joseph O. Deasy, Memorial Sloan Kettering Cancer Center, New York, NY 10065, Phone: 212-639-8413; E-mail: deasyj@mskcc.org

Clin Cancer Res 2020;26:4643–50

doi: 10.1158/1078-0432.CCR-19-2627

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

Translational Relevance

Results of the randomized controlled phase III radiotherapy trial NRG Oncology/RTOG 0617 for locally advanced non–small cell lung cancer showed an unexpected increase in mortality within 2 years following treatment in the high-dose arm. This detailed modeling study, performed according to a preregistered analysis plan, shows that differential mortality was primarily associated with dose-volume loads on multiple cardiopulmonary structures. The resulting model, applied to set-aside validation data, shows a strong ability to discriminate risk, as well as good calibration. Although the model does not identify cause, it is a quantitative tool that could be used in treatment planning to reduce mortality risk through adjustment of dose patterns outside the tumor volume.

Materials and Methods

To increase rigorousness and promote transparency, the study analysis plan (SAP) was logged on the Open Science Foundation website prior to receipt of the dataset, and is available from that site at the time of this writing (10), as well as in the online Supplementary Materials and Methods. Any departure or addition relative to the SAP is explicitly stated in the following.

Investigated cohort

All patients treated in RTOG 0617 that had complete dose-volume histogram (DVH) data through retrievable treatment plans were included in this study (437/554 patients that were initially enrolled). Characteristics for this cohort are summarized in **Table 1**. The median follow-up time across all 437 patients was 24 (range: 0.5–97) months. The original study was sponsored by the NCI and received central institutional review board approval under NCI, and the associated ethical guidelines adheres to the Belmont report. All patients read and signed informed consent documents (1).

Three heart dose-volume thresholds were recommended as treatment planning guidelines in the original trial (Heart V33% < 60 Gy; V66% < 45 Gy; and V100% < 40 Gy). These guidelines had the lowest priority among all concerned normal tissues (see Appendix 1 in ref. 1). The cardiac substructures considered (atria, pericardium, and ventricles) were based on the RTOG 1106 organ-at-risk atlas https://www.nrgoncology.org/Portals/0/Scientific Program/CIRO/Atlases/RTOG 1106 OAR Atlas for website (1).ppt (11). Segmentation was performed by five physicians following completion of the trial, as part of the analysis effort, and was reviewed by a single physician. As illustrated in the slice-by-slice definitions in (11), the pericardium started approximately 5–6 mm above the superior end of the aortic arch and ended at the diaphragm, and was the envelope of the four chambers, the aorta (primarily the ascending part), the pulmonary artery and vein, the superior vena cava, and the coronary arteries. The inferior boarder of the atria and ventricles (if still appearing; typically, only the left ventricle) were at the last slice of the pericardium, while the superior boarder was just below where the pulmonary artery passed the midline (if appearing; typically, only the left side). The left and right atria were fused and so were the left and right ventricles. Other structures analyzed included the gross tumor volume (GTV; prescribed dose and volume only), as well as the nontumor invaded lung; both obtained from the original treatment planning structure set.

Modeling

There is no universally accepted approach to dose-volume modeling, particularly given multiple potential critical tissues. The goal of our modeling strategy was to deal with the two key challenges of (i) many substructures, which increases the number of potential DVH candidate predictors, as well as (ii) a potential overreliance on predictors with maximum univariate significance, compared with other factors that could have been selected if another dataset were collected. For these reasons, we simulated model OS Cox selection variability on bootstrap. Candidate multivariate models on the basis of bootstrapped datasets were recorded. The final ensemble model was then derived by averaging the coefficients across candidate models, and then validated in the set-aside validation subset.

For the purpose of modeling, the cohort was randomly split into a training and a set-aside validation subset ($N = 306$ and $N = 131$). However, the split maintained the same fraction of patients in each prescription level arm. Furthermore, the splits were not allowed to be different (at the $P \le 0.05$ level on a Wilcoxon rank sum test) with respect to: age, systemic therapy (concurrent, consolidation, and cetuximab chemotherapy), gender, GTV, histology, lymph node group, OS status, time since randomization, performance status, prescribed dose, radiotherapy technique, smoking status, and tumor location and stage (**Table 1**; Supplementary Fig S1). This approach aimed at minimizing potential bias between the training and validation subsets because these six variables were previously reported to predict OS in (1) and to further result in a somewhat balanced split also with regards to OS status and time since randomization.

Candidate OS models

Within the training process, model building was performed using Cox proportional hazards regression based on a total of 191 variables (19 related to the disease, patient, or treatment, and 43 DVH cut-off points and volumes of each of the four structures). Dose was fractionation corrected using the linear-quadratic equation (assuming $\alpha/\beta = 3$ Gy) and was represented by the minimum and mean dose to the hottest x% volume (Dx%[Gy], MOHx%[Gy]; x was 5–100 in 5% increments; ref. 12), and the mean and max dose (Min[Gy], Mean[Gy], and Max[Gy]). This nomenclature is consistent with TG-263 (13). The minimum dose to the hottest x% volume, Dx%[Gy], variables, were used rather than VxGy[%] variables, because of preferred statistical properties: as VxGy[%] approaches the prescription dose, many patients have values of zero/close to zero, whereas this is not the case for associated Dx%[Gy].

Candidate predictors were suggested by a $P \le 0.0003$ (Bonferroni corrected for 191 variables), as stated in the SAP. If multiple candidate DVH variables were identified for each structure, the one with the lowest P value was promoted to MVA. The underlying motivation for "one best DVH variable per structure choice" was due to an anticipated strong correlation between DVH variables of the same structure, and parsimonious models are particularly preferred to avoid unnecessary model selection instability. Candidate predictors were then subject to MVA, which was conducted with forward-stepwise variable selection and a retention criterion of $P \le 0.05$ from a likelihood ratio test. Univariate analysis followed by MVA was carried out together on 1000 bootstrapped replicates (with replacement) in which MVA models (selected in ≥10% of the replicates) were considered candidate models, and were subject to validation. As was not clearly stated in the SAP, linear interaction terms between variables in the candidate MVA models were tested and incorporated into the MVA models if adhering to the univariate P value cutoff, and in addition all regression

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

Table 1. Summary of the investigated disease, patient and treatment characteristics [*n* (%) or median (range)] in the whole cohort (left), and separated between training/validation (left/right).

	All data ($N = 437$)	Training ($n = 306$)	Validation ($n = 131$)	*P*
Age (years)	64 (37–83)	64 (37–83)	64 (38–82)	0.86
Cetuximab assigned				
Yes (ref)	203 (46)	135 (44)	68 (52)	0.14
No	234 (54)	171 (56)	63 (48)	
Concurrent paclitaxel/carboplatin dose				
Other (ref)	321 (73)	227 (74)	94 (72)	0.60
85%–115%	116 (27)	79 (26)	37 (28)	
Consolidation paclitaxel/carboplatin				
No (ref)	319 (17)	55 (18)	20 (15)	0.49
Yes	362 (83)	251 (82)	111 (85)	
Gender				
Male (ref)	179 (41)	122 (40)	56 (43)	0.58
Female	258 (59)	184 (60)	75 (57)	
GTV (cm^3)	93 (4–1,194)	91 (4–1,194)	105 (6–959)	0.15
Histology				
Adeno (ref)	175 (40)	130 (42)	45 (34)	0.07
SCC (ref)	188 (43)	123 (40)	65 (50)	0.11
Large cell undifferentiated/NSCLC NOS	74 (17)	53 (17)	21 (16)	—
Lymph node group				
LLL level 7–10 (ref)	260 (59)	179 (58)	81 (62)	0.52
Other	177 (41)	127 (42)	50 (38)	
OS (months)				
Alive	118 (27)	82 (27)	36 (27)	0.33
Dead	319 (73)	224 (73)	95 (73)	
Median (95% CI) time since randomization[a]	25 (21–29)	23 (20–28)	26 (21–37)	
Prescribed dose				
60 Gy (ref)	253 (58)	176 (58)	77 (59)	0.81
74 Gy	184 (42)	130 (42)	54 (41)	
RT technique				
3DCRT (ref)	228 (52)	158 (52)	70 (53)	0.73
IMRT	209 (48)	148 (48)	61 (47)	
Smoking status				
Former (ref)	319 (73)	222 (73)	97 (74)	0.38
Current (ref)	73 (17)	48 (16)	25 (19)	0.75
Tumor inferiority				
Upper lobe (ref)	291 (67)	203 (66)	88 (67)	0.21
Lower lobe (ref)	93 (21)	70 (23)	24 (18)	0.87
Tumor laterality				
Left (ref)	171 (39)	123 (40)	69 (53)	0.46
Right (ref)	242 (55)	174 (56)	81 (62)	0.79
Tumor stage				
IIIA+N2 (ref)	293 (67)	207 (68)	86 (66)	0.68
IIIB+N3	144 (33)	100 (32)	45 (34)	
Zubrod performance				
0 (ref)	256 (59)	179 (58)	77 (59)	0.96
1	181 (41)	127 (42)	54 (41)	

Note: The *P* value is from a Wilcoxon rank sum test (exception: comparison of OS curves in which the *P* value refers to a log-rank test) between variables in training and validation.

Abbreviations: LLL, left lower lobe; ref., reference; RT, radiotherapy; SCC, squamous cell carcinoma.

[a]On the basis of Kaplan–Meier estimates.

coefficients were required to be positive given an underlying hypothesis of cardiopulmonary dose leading to worse survival.

Validation of candidate OS models.

The validation procedure was not described in detail in the SAP, but was carried out as follows as suggested by Royston and Altman (14): for each candidate MVA model, the validation procedure included calculation of the prognostic index (PI; $\beta_1 \times \text{Variable}_1 + \beta_2 \times \text{Variable}_2 \ldots$) using the coefficients obtained in the training subset. The predicted survival curve was obtained by combining the PI with the observed survival. The predicted survival curve was assessed both for calibration and discrimination (14, 15): calibration was assessed in four risk groups (low/moderate/intermediate/high risk: <16th/16th–50th/>50th–84th/>84th percentile of the predicted PI; ref. 14) with the primary focus of comparing the low- with the high-risk group, as suggested in (14). Discrimination was assessed using Harrell *c*-index,

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

which was independent of the risk group stratification (i.e., assessed using all patients in the validation subset). The model is a TRIPOD type 2b model (one dataset randomly split into a training and validation cohort; ref. 15).

Final ensemble modeling procedure

As noted above, the goal was to propose a modeling procedure that is robust with respect to dataset variability regarding predictor selection. Given the expected distribution of MVA models on bootstrap replication, that is, more than one candidate MVA model, a subsequent ensemble modeling approach was conducted. This was inspired by the approach proposed by Zhang and colleagues (16), but unlike their principal component analysis method, our generated candidate models were bagged, that is, the coefficients for each variable were averaged and weighted according to the selection frequency of the associated DVH variable. This leads to a new bagged-based PI in training; the validation procedure was analogous to that described previously.

Results

Candidate OS models in the training subset

None of the disease or patient characteristics, or structure volumes was a candidate predictor (Supplementary Fig. S2). However, dose to the four investigated structures was, and the best DVH cut-off points were Atria D45%[Gy], Pericardium MOH55%[Gy], Ventricles MOH5%[Gy], and Lung Mean[Gy] (**Fig. 1**; Supplementary Table S1). These four candidate predictors were passed on to MVA, and resulted in four candidate MVA models (model frequency given in %):

(1) Atria D45%[Gy] + Pericardium MOH55%[Gy] (18%)
(2) Atria D45%[Gy] + Pericardium MOH55%[Gy] + Ventricles MOH5%[Gy] (15%)
(3) Atria D45%[Gy] + Pericardium MOH55%[Gy] + Lung Mean[Gy] (14%)
(4) Pericardium MOH55%[Gy] + Ventricles MOH5%[Gy] + Lung Mean[Gy] (10%)

The associated *c*-index was 0.87 for MVA model 1, followed by 0.84, 0.86, and 0.84 for models 2, 3 and 4, respectively. No interaction term passed the $P \le 0.0003$ criterion (the lowest P value was observed for the interaction term between Atria D45%[Gy] and Pericardium MOH55%[Gy]; $P = 0.01$), and no such term was, thus, incorporated to any of the MVA models.

Exploration of candidate OS models in the validation subset

As expected, discrimination of the four MVA models dropped in the validation subset compared with in the training subset (*c*-index of MVA models 1/2/3/4: 0.80 vs. 0.87/0.82 vs. 0.84/0.82 vs. 0.86/0.82 vs. 0.84). Also, the PI was slightly lower in validation compared with training for all models, but not significantly so (Supplementary Fig. S3). Calibration was typically satisfactory between the low- and

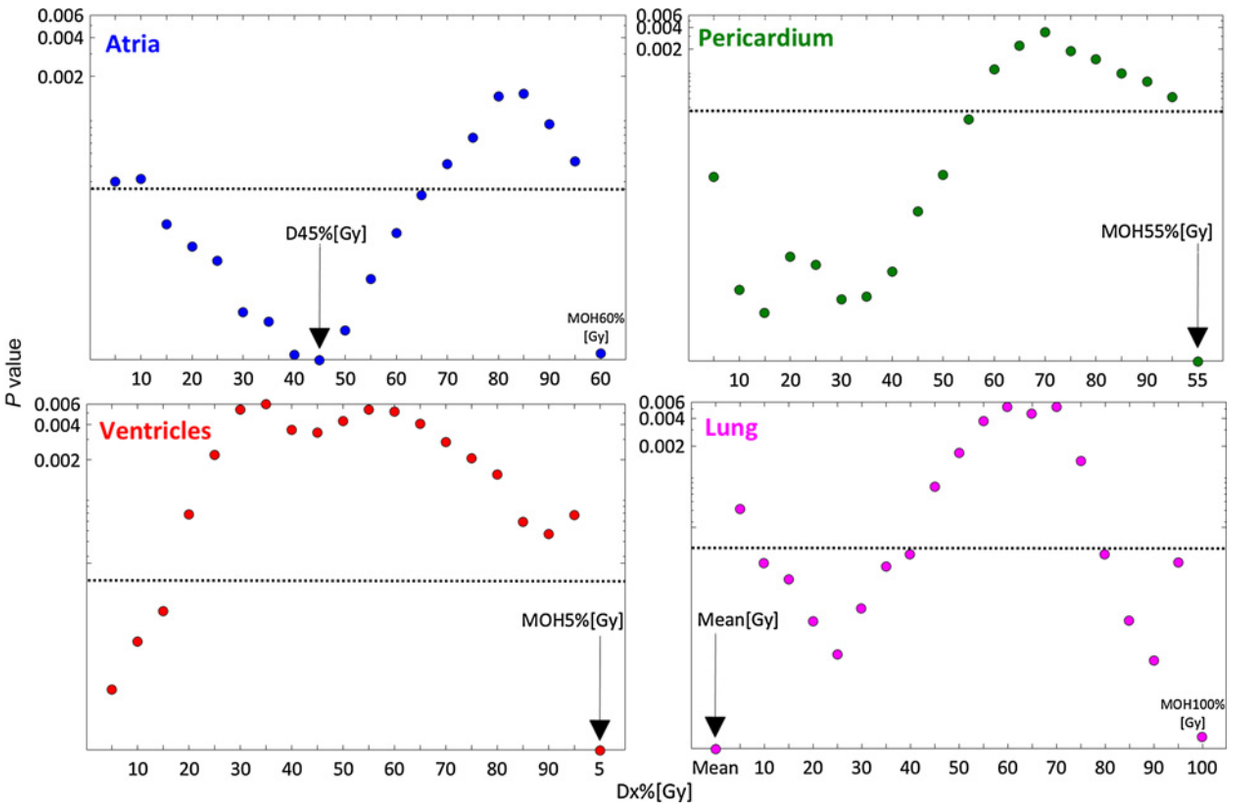

Figure 1.

Univariate *P* values (median overall samples) for the investigated Dx%[Gy] variables, and the best MOHx%[Gy] (rightmost datapoint) for each of the investigated four structures (exception: Mean[Gy] also given for lung because this was the best DVH predictor). The best DVH predictor for each structure is denoted with a black arrow, the dotted black lines represent the Bonferroni-corrected significance level at $P = 0.0003$, and the *y*-axis is on a log scale.

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

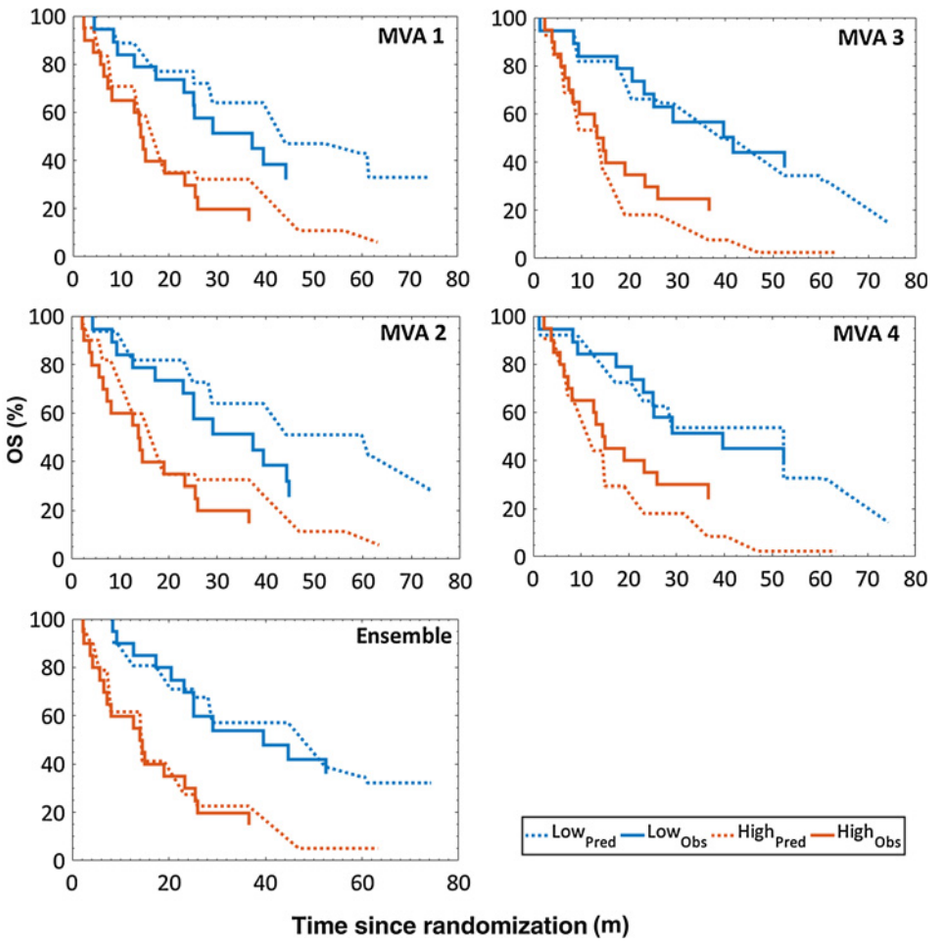

Figure 2.
Kaplan–Meier curves for the low- and high-risk groups based on MVA models 1–4 and the ensemble model comparing the observed survival rates (solid) versus the predicted survival rates (dotted) in validation; the latter modifying the observed survival curve in validation based on the PI from training. The model predicts the split of the risk curves not the shared underlying survival curve, which is fixed by the whole validation dataset in the proportional hazards model.

the high-risk groups, but somewhat inferior between the moderate- and the intermediate-risk groups (**Fig. 2**; Supplementary Fig. S4). At 18 and 36 months in the high- and the low-risk groups, the predicted survival rate was higher than the observed survival rate for MVA models 1 and 2, whereas for MVA models 3 and 4 the predicted survival rate was on average lower than the observed survival rate for the corresponding time points and risk groups (**Fig. 2**; Supplementary Fig. S4; Supplementary Table S2).

Ensemble modeling

Bagging was performed on the four final models, that is, Atria D45%[Gy], Pericardium MOH55%[Gy], Ventricles MOH5%[Gy], and Lung Mean[Gy]. The PI in validation was located between the PIs from MVAs 1–2 and MVAs 3–4 (mean: 0.89 vs. 0.63–0.66, and 1.11–1.11). The associated *c*-index in validation increased from 0.80–0.82 for the individual MVA models to 0.89 for the bagged model, and in addition, risk group stratification between the high- and the low-risk group was excellent (**Fig. 2**): for the bagged model, the predicted versus observed survival rate differed in two percentage points (78% vs. 80%) at 18 months and six percentage points (40% vs. 34%) at 36 months in the low-risk group. The equivalent difference in the high-risk group was one and three percentage points (41% vs. 40%, and 23% vs. 20%). These numbers were considerably smaller compared with the corresponding survival rates for MVA models 1–4 (Supplementary Table S2), in which the median difference was five and eight percentage points at 18 and 36 months in the low-risk group, and 17 and 15 percentage points for the same time points in the high-risk group.

In the high-risk group, and based on the ensemble model, the population average of Atria D45%[Gy], Pericardium MOH55%[Gy], Ventricles MOH5%[Gy], and Lung Mean[Gy] were 44, 51, 56, and 17 Gy (all doses given as the equivalent dose in 2 Gy fractions assuming $\alpha/\beta = 3$ Gy). For more conservative treatments, and if feasible, the upper limits for treatment planning could be defined by combining the intermediate- and the high-risk group (population average: Atria D45%[Gy] ≤ 30 Gy; Pericardium MOH55%[Gy] ≤ 39 Gy; Ventricles MOH5%[Gy] ≤ 41 Gy; and Lung Mean[Gy] ≤ 15 Gy). In addition, Supplementary Fig. S5 provides an overview of the predicted and observed OS using the ensemble model (within2 years postrandomization, which was the cohort median follow-up time for OS and also the follow-up time studied in further detail in ref. 1). Lower dose-volume values would of course be preferred if feasible.

Furthermore, the coefficients behind the PI for the ensemble model ($PI_{Ensemble}$: 0.02 × Pericardium MOH55%[Gy] + 0.002 × Atria D45%[Gy] + 0.002 × Ventricles MOH5%[Gy] + 0.03 × Lung Mean[Gy]), and the coefficients for all other univariate and multivariate models are given in Supplementary Table S1. The ensemble PI can be used to construct predicted survival given the presumed base survival curve (alternatively one can use the observed survival provided here if assumed to be similar across cohorts) in external data, for example, for validation purposes adhering to the procedures used here and as given in more detail in (14).

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

Discussion

Given the excellent discrimination of survival time based on predictions from the bagged model in the set-aside validation subset ($c = 0.89$), and the use of multi-institutional data from a randomized controlled phase III trial allowing for limited correlations between the investigated variables, this model gives further insights into the unexpected OS findings of RTOG 0617. This model emphasizes associations between dose to the major blood-carrying structures and OS. Other treatment-related factors such as GTV, prescription dose, and radiotherapy technique did not pass even the univariate stage of modeling ($P = 0.10$, 0.27, and 0.39).

The bagged model combined the models that emerged on bootstrap: dose to the atria, lung, pericardium, and ventricles. Previously published studies have predominantly suggested either that dose to cardiac structures (4–6), and/or dose to the lung (3) predicts OS for this patient group. In agreement with the two previously published OS models based on the same RTOG 0617 data (1, 17), which included heart dose (1), all four candidate MVA models suggested here include dose to the pericardium (through MOH55%[Gy]). The pericardium in this study includes the four chambers, the aorta (primarily the ascending part), the pulmonary artery and vein, the superior vena cava, and the coronary arteries (11). Also, frequently selected was the Atria D45%[Gy], present in three of the four candidate MVA models together with Pericardium MOH55%[Gy], or with/without Lung Mean[Gy], or Ventricles MOH5%[Gy]. Ventricles MOH5%[Gy] was the only DVH predictor that was related to the highest doses. While the model suggests striving for dose limits given in the Results section, the ranges of the dose-volume parameters among those patients who survived is wide (Atria D45%[Gy]: 0.2–61 Gy, Pericardium MOH55%[Gy]: 0.6–58 Gy, Ventricles MOH5%[Gy]: 0.4–38 Gy, and Lung Mean[Gy]: 1.5–17 Gy), which indicates the possibility for introducing dose-volume constraints in general also in light of the low priority on dose sparing to the heart in the original 0617 trial (refer Appendix 1 in ref. 1).

Others have recently reported associations between dose-volume metrics and OS following radiotherapy in patients with LA-NSCLC based on primarily single-institutional data from nonrandomized trials. Vivekanden and colleagues (6) found that left atrial V63–69Gy[%] predicted OS, and was the only variable significantly doing so in their substructure MVA model ($N = 78$). Stam and colleagues (4) reported an association between OS and heart V2Gy[%] using a contouring-free dose mapping approach focusing only on the heart and lungs ($N = 469$). McWilliam and colleagues (5) also used a "contouring-free" approach ($N = 1{,}101$), and established an association between OS and dose to the base of the heart (aorta, the region of origin of the coronary arteries, and the sinoatrial node). While those three studies found that cardiac dose increases the risk of death, Tucker and colleagues (3) established a similar association ($N = 468$), but instead using Lung Mean[Gy]. However, it should be pointed out that doses to heart substructures were not analyzed. In contrast to (3–8), this study was based on data from multiple institutions, which includes an inherently broad variability in DVH variables, the modeling approach carefully followed the validation procedures outlined in the anchor publication by Royston and Altman (14), and further also adhered to the TRIPOD statement (15). In addition, the SAP associated with this study was preregistered on a public domain, which allows for analysis transparency.

Perhaps most relevant to our analysis is the OS analysis by Speirs and colleagues (ref. 8; $N = 416$), that includes DVH metrics of both the cardiac and the pulmonary system, resulting in significant variables of heart V50Gy[%], heart volume, and lung V5Gy[%], which is related to our result. Another closely related result is that of Contreras and colleagues (7), who derived a multivariable Cox model showing that heart V50Gy[%] together with gender, and an elevated blood neutrophil to lymphocyte ratio 4 months after radiotherapy was associated with OS ($N = 400$). They did, however, not exploit dose to any other parts of the blood-carrying tract, and the model was not validated in set-aside data. Taken together, these recent publications, including also this study, emphasize a role of the heart, lungs, and large vessel irradiation on OS.

As stated in Methods and Materials, only one DVH variable per structure was passed on to MVA assuming a strong anticipated correlation between DVH variables of the same structure. The intrastructural correlation between each final candidate DVH metric and the other DVH metrics that passed the univariate P value cutoff (31/43 metrics for atria, lung, and pericardium, and 22/43 ventricle metrics) for that structure was indeed strong (median Spearman rank correlation coefficient, $R_s = 0.97$) for both Atria D45%[Gy] and Pericardium MOH55%[Gy], but was somewhat weaker for Ventricles MOH5%[Gy] and Lung Mean[Gy] (median $R_s = 0.93$ and 0.89). For instance, the correlation with Mean[Gy] was typically among the strongest for Atria D45%[Gy], Pericardium MOH55%[Gy], and Ventricles MOH5%[Gy] ($R_s = 0.97$, 0.99, and 0.91). The interstructural correlation between the four final predictors was lower than that of the intrastructural correlation (median $R_s = 0.69$), but not surprisingly of the same magnitude between Pericardium MOH55%[Gy], and either Atria D45%[Gy], or Ventricles MOH5%[Gy] ($R_s = 0.85$, 0.82). In summary, correlations were strong throughout most of the investigated dose range. For illustration purposes, which was not specified on the SAP, for each structure the univariate DVH predictor that presented with the weakest correlation with the final predictor for each structure was added to the ensemble model in turn [Atria MOH15%[Gy], Pericardium D5%[Gy], and Lung D5%[Gy]: $R_s = 0.80$, 0.63, and 0.48 with the final predictor for each structure; note: no ventricles DVH variable was included because ventricles MOH5%[Gy] was strongly correlated with all ventricle variables (minimum $R_s = 0.90$)]. A gain in c-index, as compared with that of the ensemble model, was not observed. A last note regarding correlation: even though the strongest correlations were accounted for, some degree of collinearity will inevitably remain given that the four DVH variables in the final model were derived from the similar anatomic region.

A higher cardiopulmonary dose being associated with OS does not necessarily translate into successful estimation of a causal role in mortality, and we would like to emphasize that this study does not attempt to propose an underlying causal factor. While such a link has been established between heart irradiation and cardiac dysfunction in large breast cancer and Hodgkin lymphoma series (18, 19), a similar dose-response for NSCLC has, thus far, only been suggested in two small cohort studies ($N = 125$ and 112; refs. 20, 21). Of the eight available studies (including also this study) that have analyzed OS after radiotherapy for LA-NSCLC and have estimated a relationship with dose to the cardiopulmonary system (1, 3–8), only in (6) and in (7) was an effort made to untie causality from correlation. The approach taken in (6) was to assess cardiac status through ECG measurements, and define abnormality as any new ECG event 6 months after completed radiotherapy relative to baseline. Two MVA models were generated; one based on heart dose, and one based on heart dose as well as dose to cardiac substructures. Heart dose and ECG changes constituted the former MVA model, while the latter MVA model included only left atrial dose. A plausible explanation could, for example, be that ECG changes are associated with damage caused by dose to left atrial and, thus, left atrial dose being a proxy for cardiac toxicity, and, therefore,

ECG changes did not remain significant. However, an explanation to this was not provided, and given that the analyzed cohort was limited in size in particular for the analysis regarding ECG changes (53/78 patients; Table 2 legend in ref. 6), this should be explored in an independent, and ideally larger cohort. Another attempt to untangle causality was made in the study by Contreras and colleagues (7) and in the study by Thor and colleagues (22) who found indications of an unprecedented immune suppression explaining OS. Thus far, however, no study has simultaneously explored cardiopulmonary function and immune suppression in the setting of OS. In summary, the causal etiology of our established dose–response model remains to be further elucidated including, but not limited to, whether it is in the direction of cardiac dysfunction, immunosuppression, or both.

None of the investigated non-DVH characteristics were candidate predictors for OS. This included also radiotherapy technique ($P = 0.39$) and prescription dose level ($P = 0.27$). A preceding analysis based on the 0617 data neither found an association between radiotherapy technique and OS (23). Also in the same data, prescription dose level was previously found to be a predictor (1), and although this variable was included in the MVA in (1), the original model, based on the 0617 data was dominated by Heart V5Gy[%] ($P = 0.02$ vs. 0.004; refer Appendix 6 in ref. 1). Of note, even in the entire cohort investigated in this study the effect of prescription dose level on OS was diminished as opposed to in (ref. 1; $P = 0.02$ vs. 0.004; bottom Supplementary Fig. S6 vs. **Fig. 2** in ref. 1). Furthermore, in splitting the training and the validation subsets, the prescription dose level, OS status, and time since randomization were kept as similar as possible and were at least not significantly different. As shown in top Supplementary Fig. S6, this splitting approach did not guarantee a preserved pattern in prescription dose level. Although with considerably smaller tumors than in this study (median: 69 cm^3, 35 cm^3 vs. 93 cm^3), the MVA models suggested in (4) and in (5) included tumor volume ($P = 0.006$ and $P < 0.001$). Tumor volume was also in the MVA model in (ref. 3; median: 128 cm^3; $P < 0.0001$). Age was not a predictor here ($P = 0.27$), even though the age distribution was similar across this study and [ref. 4; median (range): 65 (30–87) years vs. 64 (37–83) years] in which the MVA model included age ($P = 0.001$). Age was also present in the MVA model suggested in [ref. 5; median (range): 73 (38–95) years; $P = 0.04$]. Lymph node involvement in the left lower lobe, and lower/middle primary tumors was not found to predict OS in this study ($P = 0.001$ and 0.02). In the study by Speirs and colleagues (8), bilateral mediastinal lymph node involvement was present in their MVA model ($P = 0.03$), and this was due to tumors being located in the left lower lobe. The MVA model by McWilliam and colleagues (5) included nodal stage ($P = 0.01$).

In conclusion, based on the multi-institutional randomized RTOG 0167 data, we showed that dose to the cardiopulmonary system (rather than dose to isolated components of this system or other characteristics), defined quantitatively by the model, compromises OS after radiotherapy for LA-NSCLC. While the causal etiology of this effect remains to be further elucidated, the possibility to minimize risk using the model DVH metrics could be explored in standard treatment planning systems.

Disclosure of Potential Conflicts of Interest

J.O. Deasy reports receiving commercial research grants from Varian Oncology Solutions, Philips Healthcare Corp., and Elekta, and holds ownership interest in PAIGE.AI Inc. C. Hu is an employee/paid consultant for Merck & Co. C. Robinson is an employee/paid consultant for Varian, AstraZeneca, and Cardialogica, and reports receiving commercial research grants from Varian. J. Bradley is an advisory board member/unpaid consultant for AstraZeneca, Inc. and Mevion Medical Systems, Inc. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: M. Thor, J.O. Deasy, C. Hu, E. Gore, V. Bar-Ad, C. Robinson
Development of methodology: M. Thor, J.O. Deasy, C. Hu, V. Bar-Ad, C. Robinson, M. Wheatley, J.H. Oh, Y. Xiao
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.O. Deasy, C. Hu, E. Gore, V. Bar-Ad, C. Robinson, M. Wheatley, J. Bogart, Y.I. Garces, V.S. Kavadi, P. Iyengar, J.S. Witt, J.W. Welsh, J.M. Larner, Y. Xiao, J. Bradley
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Thor, J.O. Deasy, C. Hu, C. Robinson, J.H. Oh, Y.I. Garces, P. Iyengar, J. Bradley
Writing, review, and/or revision of the manuscript: M. Thor, J.O. Deasy, C. Hu, E. Gore, V. Bar-Ad, C. Robinson, J.H. Oh, J. Bogart, Y.I. Garces, V.S. Kavadi, S. Narayan, P. Iyengar, J.S. Witt, C.D. Koprowski, Y. Xiao
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.O. Deasy, C. Hu, J.W. Welsh, J. Bradley
Study supervision: J. Bradley
Other (principal investigator of the primary RTOG 0617 article, and reviewed tumor and organs-at-risk contours for the trial patients): J. Bradley

Acknowledgments

This study was funded through the NIH/NCI R01 CA198121 (PI, J.O. Deasy) and the NIH/NCI Cancer Center Support Grant P30 CA008748.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 9, 2019; revised February 7, 2020; accepted May 7, 2020; published first May 12, 2020.

References

1. Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. Lancet Oncol 2015;16:187–99.
2. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Durvulumab after chemoradiotherapy in stage III non-small cell lung cancer. N Engl J Med 2017;377:1919–29.
3. Tucker SL, Liu A, Gomez D, Tang LL, Allen P, Yang J, et al. Impact of heart and lung dose on early survival in patients with non-small cell lung cancer treated with chemoradiation. Radiother Oncol 2016;119:495–500.
4. Stam B, van der Bijl E, van Diessen J, Rossi MMG, Tijhuis A, Belderbos JSA, et al. Heart dose associated with overall survival in locally advanced NSCLC patients treated with hypofractionated chemoradiotherapy. Radiother Oncol 2017;125:62–5.
5. McWilliam A, Hodgson C, Vasquez Osario E, Faivre-Finn C, van Herk M. Radiation dose to heart base linked with poorer survival in lung cancer patients. Eur J Cancer 2017;85:106–13.
6. Vivekanden S, Landau DB, Counsell N, Warren DR, Khwanda A, Rosen SD, et al. The impact of cardiac radiation dosimetry on survival after radiation therapy for non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2017;99:51–60.
7. Contreras JA, Lin AJ, Weiner A, Speirs C, Samson P, Mullen D, et al. Cardiac dose is associated with immunosuppression and poor survival in locally advanced non-small cell lung cancer. Radiother Oncol 2018;128:498–504.
8. Speirs CK, DeWees TA, Rehman S, Molotievschi A, Velez MA, Mullen D, et al. Heart dose in an independent dosimetric predictor of overall survival in locally advanced non-small cell lung cancer. J Thorac Oncol 2017;12:293–301.

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

9. Deasy JO, Bentzen SM, Jackson A, Ten Haken RK, Yorke ED, Constine LS, et al. Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. Int J Radiat Oncol Biol Phys 2010;76:151–4.
10. Thor M, Deasy JO. The role of heart-related dose-volume metrics on overall survival in the RTOG 0617 clinical trial. Available from: https://doi.org/10.17605/OSF.IO/HZSVA.
11. Wheatley MD, Gore EM, Bar Ad V, Robinson CG, Bradley JD. Defining a novel cardiac contouring atlas for NSCLC using cadaveric anatomy. Int J Radiat Oncol Biol Phys 2014;90:658.
12. El Naqa I, Suneja G, Lindsay PE, Hope AJ, Alaly JR, Vicic M, et al. Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. Phys Med Biol 2006; 51:5719–35.
13. Mayo CS, Moran JM, Bosch W, Xiao Y, McNutt T, Popple R, et al. American Association of Physicists in Medicine Task Group 263: standardizing nomenclatures in radiation oncology. Int J Radiat Oncol Biol Phys 2018;100:1057–66.
14. Royston T, Altman DG.External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;33:1–15.
15. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariate prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:1–73.
16. Zhang J. Developing robust non-linear models through bootstrap aggregated neural networks. Neurocomputing 1999;25:93–113.
17. Bradley JD, Hu C, Komaki RR, Masters GA, Blumenschein GR, Schild SE, et al. Long-term of NRG Oncology RTOG 0617: standard- versus high-dose chemoradiotherapy with or without cetuximab for unresectable stage III non-small cell lung cancer. J Clin Oncol 2020;38:706–14.
18. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med. 2013;368:987-98.
19. van Nimwegen FA, Schaapveld M, Cutter DJ, Janus CP, Krol AD, Hauptmann M, et al. Radiation dose-response relationship for risk of coronary heart disease in survivors of Hodgkin lymphoma. J Clin Oncol 2016;34:235–43.
20. Dess RT, Sun Y, Matuszak MM, Sun G, Soni PD, Bazzi L, et al. Cardiac events after radiation therapy: combined analysis of prospective multicenter trials for locally advanced non- small-cell lung cancer. J Clin Oncol 2017;35:1395–402.
21. Wang K, Eblan MJ, Deal AM, Lipner M, Zagar TM, Wang Y, et al. Cardiac toxicity after radiotherapy for stage III non-small-cell lung cancer: pooled analysis of dose-escalation trials delivering 70 to 90 Gy. J Clin Oncol. 2017; 35:1387–94.
22. Thor M, Montovano M, Hotca A, Luo L, Jackson A, Wu AJ, et al. Are unsatisfactory outcomes after concurrent chemoradiotherapy for locally advanced non-small cell lung cancer due to a treatment-related immunosuppression. Radiother Oncol 2020;143:51–57.
23. Chun SG, Hu C, Choy H, Timmerman RD, Schild SE, Bogart JA, et al. Impact of intensity-modulated radiation therapy technique for locally advanced non-small-cell lung cancer: a secondary analysis of the NRG Oncology RTOG 0617 randomized clinical trial. J Clin Oncol 2017;35:56–61.

Published OnlineFirst May 12, 2020; DOI: 10.1158/1078-0432.CCR-19-2627

Clinical Cancer Research

Modeling the Impact of Cardiopulmonary Irradiation on Overall Survival in NRG Oncology Trial RTOG 0617

Maria Thor, Joseph O. Deasy, Chen Hu, et al.

Clin Cancer Res 2020;26:4643-4650. Published OnlineFirst May 12, 2020.