Original Article

Multifactorial risk factors for mortality after chemotherapy and radiotherapy for non-small cell lung cancer

Gilles Defraene [a,*], Frank J.W.M. Dankers [b,c], Gareth Price [d], Ewoud Schuit [e], Wouter van Elmpt [c], Soumia Arredouani [a], Maarten Lambrecht [a], Joost Nuyttens [f], Corinne Faivre-Finn [d], Dirk De Ruysscher [a,c]

[a] KU Leuven – University of Leuven, Department of Oncology, Experimental Radiation Oncology, Leuven, Belgium; [b] Radboud University Medical Center, Department of Radiation Oncology, Nijmegen; [c] Maastricht University Medical Center, Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+, The Netherlands; [d] Division of Cancer Sciences, The University of Manchester, The Christie NHS Foundation Trust, Manchester, UK; [e] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University; and [f] Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

ARTICLE INFO

ABSTRACT

Background and purpose: A higher radiation dose to the heart is known to be associated with increased mortality in non-small cell lung cancer (NSCLC) patients. It is however unknown what the contribution of the heart dose is when other risk factors for mortality are also accounted for.

Materials and methods: We constructed and externally validated prediction models of mortality after definitive chemoradiotherapy for NSCLC. Models were developed in 145 stage I-IIIB NSCLC patients. Clinical (performance status, age, gross tumour volume (GTV) combining primary tumour and involved lymph nodes, current smoker) and dosimetric (mean lung (MLD) and heart (MHD) dose) variables were considered. Multivariable logistic regression models predicting 12 and 24 month mortality were built in 5-fold cross-validation. Discrimination and calibration was assessed in 3 external validation datasets containing 878 (via distributed learning), 127 and 96 NSCLC patients.

Results: The best discriminating prediction models combined GTV, smoker and/or MHD: bootstrapping AUC (95% CI) of 0.74 (0.66–0.78) and 0.69 (0.55–0.74) at 12 and 24 months. At external validation, the 24 month mortality GTV-smoker-MHD model robustly showed moderate discrimination (AUC = 0.61–0.64 before and 0.64–0.65 after model update) with limited 0.01–0.07 improvement over a GTV-only model, and calibration slope (0.64–0.65). This model can identify patients for whom a MHD reduction may be useful (e.g. PPV = 77%, NPV = 52% (60% cut-off)).

Conclusions: Tumour volume is strongly related to mortality risk in the first 2 years after chemoradiotherapy for NSCLC. Modelling indicates that efforts to reduce cardiac dose may be relevant for small tumours and that smoking has an important negative association with survival.

© 2019 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 152 (2020) 117–125

Locally advanced non-small cell lung cancer (LA-NSCLC) when radically treated with standard concurrent chemoradiotherapy has a high local recurrence rate of approximately 30% at 2 years [1,2]. Additionally, the overall survival rates are low, around 30% in selected patients at 5 years in recent clinical trials investigating modern chemoradiotherapy [3–5]. Treatment-induced toxicity may partly be responsible for these unfavorable results, as indicated by the higher mortality in the high-dose arm of RTOG0617 [2,3]. A secondary analysis of this study showed the volume of the heart receiving >50 Gy to be significantly associated with overall survival in multivariable analysis [6].

To estimate the toxicity risk, population-based models of adverse events are used, for example a mean lung dose (MLD) planning constraint of 20 Gy is recommended to limit the rate of radiation pneumonitis [7]. These constraints are however not well defined at present, especially for the heart dose.

There is a growing body of evidence supporting the hypothesis that radiation to the heart induces cardiac injury and non-cancer-related deaths in lung cancer patients. Although the independent association of heart dose with mortality could not be replicated in two other studies, several associations corroborating the RTOG0617 findings have been published recently. In LA-NSCLC it was shown that the mean heart dose was associated with grade ≥3 [10] or symptomatic cardiac events [11]. In patients with stage I–II NSCLC treated with stereotactic body radiotherapy, the dose to the base of the heart was associated with non-cancer death [12]. Analyses of large routine NSCLC radiotherapy planning and image

* Corresponding author at: KU Leuven – University of Leuven, Department of Oncology, Experimental Radiation Oncology, B-3000 Leuven, Belgium.

E-mail address: gilles.defraene@uzleuven.be (G. Defraene).

guidance datasets have identified a region at the base (i.e., superioposterior region) of the heart that significantly was associated with reduced survival [13]. Furthermore a recent population-based study showed that set-up errors that shift the high dose region towards the heart correlate with excess mortality compared to those that move dose away from the heart [14].

The accurate estimation of the individual mortality risk including the impact of heart dose would allow personalization of planning dose constraints. This could facilitate the selection of high-risk patients that may benefit from emerging radiotherapy technologies, enabling high precision delivery of the radiation dose (i.e., lowering the heart dose) such as proton therapy and the magnetic resonance-guided linear accelerator radiotherapy (MR-linac) [15]. Published externally validated prediction models of lung cancer mortality focused on clinical and treatment-related risk factors and did not include dosimetric factors for heart or lungs [16,17]. These studies reported moderate model accuracy when validated externally with area under the curve around 0.60. Others studied the influence of the heart dose on mortality in a monocentric dataset, but did not include important risk factors such as tumour volume [6,9] or did not report prediction model parameters [6,8,9,13]. Moreover, modelling was mostly not repeated for different time points after treatment. In order to show the advantage of dose reductions to organs at risk with emerging technologies, based on the predicted mortality risk-driven patient selection in a realistic randomized setting, it is crucial to choose an optimal (as early as possible) mortality endpoint during model development.

In this work, we constructed and externally validated logistic regression mortality prediction models at 1 and 2 year after treatment in radically treated NSCLC cohorts from 4 separate institutions. We investigated a combined set of clinical and dosimetric risk factors, including tumour volume and dose to heart and lungs. The additional benefit of dosimetric factors for the performance of prediction models was studied in terms of model discrimination and calibration.

Materials and methods

Datasets

Stage I-IIIB NSCLC patients treated with radical chemoradiotherapy at MAASTRO Clinic during two time periods (previously collected datasets from 2003–2006 and 2014–2016 periods) constituted the development dataset. We excluded patients who did not receive chemotherapy either before or during radiotherapy, patients treated with stereotactic body radiotherapy, and patients who had received previous radiotherapy to the thorax. One hundred forty five patients with complete data were included. Prescription doses varied during the two time periods and included 66 Gy (2.75 Gy fractions with sequential chemotherapy or 2 Gy fractions with concurrent chemotherapy), 72 Gy (1.8 Gy fractions twice daily), 45 Gy (1.5 Gy fractions twice daily) followed by up to 24 Gy (2 Gy fractions), and isotoxically dose-escalated radiotherapy up to 106.4 Gy (24 fractions, positron emission tomography (PET)-boost randomized study [18]). Treatment dose was calculated on a free-breathing or mid-ventilation CT-scan [19] with convolution-superposition or Varian's Acuros algorithms, using 3D-conformal radiotherapy (3D-CRT), volumetric modulated arc therapy (VMAT) or hybrid VMAT [20] techniques. Margins of 5 mm from gross tumour volume (GTV) or internal target volume (ITV) to the clinical target volume (CTV) and 5–10 mm from CTV to the planning target volume (PTV) were applied.

The first external validation dataset consisted of 878 NSCLC patients radically treated with 3D-CRT, intensity-modulated radiotherapy (IMRT) or VMAT (chemo)radiotherapy at The Christie NHS Foundation Trust between 2005 and 2017 (ethical approval ref. 17/

NW/0060). Prescription doses were 55 Gy (2.75 Gy fractions) with radiotherapy alone or with sequential chemotherapy, and 60–66 Gy (2 Gy fractions) with concurrent chemotherapy or radiotherapy alone. In patients planned using 4D CT, the motion adjusted GTV was contoured on the Maximum Intensity Projection (MIP) image. GTV was then recovered through the method of Johnson et al. [21]. All patients were planned using the Philips Pinnacle treatment planning system with treatment margin protocols as described above for the MAASTRO cohort.

A second external validation dataset consisted of 127 NSCLC patients from Erasmus MC Rotterdam treated with radical (chemo)radiotherapy between 2009 and 2013. Prescription doses were 66 Gy (2 Gy fractions) for concurrent chemoradiotherapy and 45–60 Gy (3 Gy fractions) for sequential chemoradiotherapy. Planning was done with 3D-CRT. Margins of 5 mm were used from primary tumour GTV to CTV. As the hilar and mediastinal lymph nodes were contoured as nodal CTV, nodal GTV had to be estimated based on the primary tumour GTV-CTV association.

The third external validation dataset consisted of 96 NSCLC patients from UZ Leuven treated with radical chemoradiotherapy (mostly 66 Gy in 2 Gy fractions with concurrent chemotherapy) with 3D-CRT or IMRT and margins as for the MAASTRO cohort, between 2011 and 2016. The institutional review boards of all centers have approved the study.

Clinical and dosimetric variables were chosen based on their importance as prognostic factors for overall survival in the literature: baseline World Health Organization performance status [8,13,16,17], age at start of treatment [16], current smoker at time of diagnosis (yes/no) [22], the available dosimetric variables mean lung dose (based on both lungs) [6,8] and mean heart dose (MHD) [6,9], and GTV combining primary tumour and involved lymph nodes volumes [8,13,17,23,24]. GTV was chosen instead of TNM stage as it was previously shown to be a more significant prognostic factor for overall survival [17,23,24]. The heart was delineated in all datasets along with the pericardial sac from its most caudal part at the apex up to the beginning of the large vessels cranially.

Logistic and Cox regression model development

Survival times were calculated from the last day of radiotherapy treatment, except for the Christie cohort where this was the first day of radiotherapy. In the development dataset, Kaplan–Meier analyses assessed the association of all variables with survival. Survival curves were generated with continuous variables grouped with respect to their median value. Logistic regression models were built with the endpoints of 12 month and 24 month mortality. Censored observations at these time points (20 cases at 12 months and 25 cases at 24 months) were discarded. There were no missing data in the development dataset and non-linear transformations (log, square root, inverse transformations, etc.) of continuous variables were tested in the univariable analyses. Backward stepwise model building processes (including all variables) were followed based on the Akaike's Information Criterion which incorporates a penalty on the number of model parameters to avoid overfitting [25]. Based on one hundred times repeated 5-fold cross-validation (model building performed in every fold), the most frequently built models were selected for further analysis. Final model coefficients were determined by fitting these selected models on the complete development dataset. Cox regression models were generated for reference using the same procedure.

Model discrimination was assessed by the area under the receiver operating characteristic curve (AUC) or c-statistic. Model calibration was assessed by calibration plots correlating predicted mortality probabilities and observed mortality. Internal validation was performed by repeating the modelling within 500 bootstrap samples that were of equal size as the study population and were

drawn with replacement. This resulted in a shrinkage factor to be multiplied with the regression coefficients. Using the shrunk coefficients should give more generalizable predictions out of sample.

An AUC corrected for optimism was calculated by averaging the optimism (obtained when applying bootstrap sample-derived model coefficients to predict risks in the whole dataset) over all

Table 1
Patient and treatment characteristics in the development dataset and the 3 external validation datasets. Median and range or absolute numbers and proportions.

	Development set	External validation sets		
	MAASTRO 145 patients (2003–2016)	The Christie 878 patients (2005–2017)	Erasmus MC 127 patients (2009–2013)	UZ Leuven 96 patients (2011–2016)
Age (years)	65 (39–88)	71 (32–93)	62 (30–80)	64 (42–89)
Tumour T stage				
1	13 (9.0 %)	102 (11.6 %)	19 (15.0 %)	11 (11.5 %)
2	42 (29.0 %)	290 (33.0 %)	20 (15.8 %)	9 (9.4 %)
3	14 (9.7 %)	241 (27.5 %)	25 (19.7 %)	24 (25.0 %)
4	58 (40.0 %)	204 (23.2 %)	59 (46.5 %)	28 (29.2 %)
Unknown	18 (12.4 %)	41 (4.7 %)	4 (3.2 %)	24 (25.0 %)
N stage				
0	31 (21.4 %)	272 (31.0 %)	20 (15.8 %)	12 (12.5 %)
1	2 (1.4 %)	114 (13.0 %)	10 (7.9 %)	9 (9.4 %)
2	55 (37.9 %)	298 (33.9 %)	79 (62.2 %)	37 (38.5 %)
3	40 (27.6 %)	165 (18.8 %)	18 (14.2 %)	20 (20.8 %)
Unknown	17 (11.7 %)	29 (3.3 %)	NA	18 (18.8 %)
Histology				
Squamous cell carcinoma	50 (34.5 %)	418 (47.6 %)	46 (36.2 %)	28 (29.2 %)
Adenocarcinoma	31 (21.4 %)	223 (25.4 %)	49 (38.6 %)	41 (42.7 %)
NSCLC NOS	16 (11.0 %)	94 (10.7 %)	NA	2 (2.1 %)
Large cell	37 (25.5 %)	7 (0.8 %)	32 (25.2 %)	8 (8.3 %)
Other	3 (2.1 %)	31 (3.5 %)	NA	NA
Unknown	8 (5.5 %)	105 (12.0 %)	NA	17 (17.7 %)
WHO performance status at baseline				
0	51 (35.2 %)	89 (10.1 %)	42 (33.1 %)	10 (10.4 %)
1	78 (53.8 %)	379 (43.2 %)	73 (57.5 %)	49 (51.0 %)
2	15 (10.3 %)	293 (33.4 %)	5 (3.9 %)	1 (1.0 %)
≥3	1 (0.7 %)	84 (9.6 %)	NA	NA
Unknown	NA	33 (3.8 %)	7 (5.5 %)	NA
Smoking status at diagnosis				
Current smoker	53 (36.6 %)	228 (26.0 %)	118 (92.9 %) *	44 (45.8 %)
Stopped/never smoker	92 (63.5 %)	418 (47.6 %)	6 (4.7 %)**	52 (54.2 %)
Unknown	NA	232 (26.4 %)	3 (2.4 %)	NA
Chemotherapy treatment				
Concurrent	59 (40.7 %)	182 (20.7 %)	111 (87.4 %)	70 (72.9 %)
Sequential	86 (59.3 %)	143 (16.3 %)	14 (11.0 %)	14 (14.6 %)
RT alone or unknown	NA	553 (63.0 %)	2 (1.6 %)	12 (12.5 %)
Treatment technique				
3D-CRT	90 (62.1 %)	280 (31.9 %)	127 (100.0 %)	50 (52.1 %)
IMRT	NA	583 (66.4 %)	NA	43 (44.8 %)
VMAT	45 (31.0 %)	NA	NA	2 (2.1 %)
Hybrid VMAT	10 (6.9 %)	NA	NA	NA
Unknown	NA	15 (1.7 %)	NA	1 (1.0 %)
GTV volume (cc)	79.4 (0.3–996.8)	45.5 (0.3–501.5)	84.9 (0.5–510.2)	79.7 (4.2–559.4)
Dose per fraction (Gy)				
1.5	26 (17.9 %)	NA	NA	NA
1.8	37 (25.5 %)	NA	1 (0.8 %)	4 (4.2 %)
2	69 (47.6 %)	182 (20.7 %)	115 (90.6 %)	68 (70.8 %)
>2	13 (9.0 %)	696 (79.3 %)	11 (8.7 %)	23 (24.0 %)
Unknown	NA	NA	NA	1 (1.0 %)
Mean Lung Dose (Gy physical dose)	15.6 (2.7–23.6)	13.8 (3.3–25.7) (missing for N = 83)	15.6 (2.3–22.2)	16.6 (3.2–23.7)
Mean Heart Dose (Gy physical dose)	7.7 (0.1–45.2)	13.1 (0.0–35.4)	10.9 (0.0–46.0)	10.0 (0.1–41.6)
1 year survival probability (%)	61.3	62.0	78.7	81.2
<2009	52.8	NA***	NA	NA
≥2009	76.9	62.0	78.7	81.2
2 year survival probability (%)	38.7	35.1	48.0	54.4
<2009	29.2	NA***	NA	NA
≥2009	61.5	35.1	48.0	54.4
Median survival time (month)	18.1	16.5	22.6	29.0
<2009	13.9	NA***	NA	NA
≥2009	>25.2	16.5	22.6	29.0
Median follow-up time (month)	35.5	39.1	67.0	44.1
<2009	102.6	NA***	NA	NA
≥2009	22.6	39.1	67.0	44.1

Abbreviations: NSCLC NOS: Non-small cell lung cancer Not otherwise specified; WHO: World Health Organization; 3D-CRT: 3 Dimensional Conformal Radiotherapy, IMRT: Intensity-Modulated Radiotherapy; VMAT: Volumetric Modulated Arc Therapy; GTV: Gross Tumour Volume.
*Ever smoker.
**Never smoker.
***Only 5 patients treated <2009.

500 bootstrap replications and subtracting this average optimism from the development sample performance metric [26].

External model validation

The developed models were validated in 3 external validation datasets by applying the model coefficients to each individual to predict mortality risks. No censored observations were present at 12 months, while seventy four, zero and nine patients were discarded for the 24 month mortality endpoint in the Christie, Erasmus MC and UZ Leuven validation datasets, respectively. Discrimination was assessed and calibration plots were analyzed for their intercept (ideally 0, with negative and positive values indicating systematic over- and underestimation of the risks, respectively) and slope (values below and above 1 indicating model over- and underfitting, respectively) [27]. The appropriate model updating technique was chosen using a closed testing procedure, i.e., adjusting only the intercept as a baseline risk correction was compared to a slope adjustment and a model revision [28]. The validation in the Christie cohort was undertaken using the Varian Learning Portal (VLP, Varian Medical Systems, Palo Alto, CA) distributed learning platform. Distributed learning addresses information governance concerns with the traditional data sharing model in which datasets are physically centralized at one institution for analysis. Instead the analysis algorithm is sent to each site holding data, analyzing it locally and returning only the results.

Distributed learning whilst a relatively novel approach, has previously been used in several projects at the authors' institutions [16,29,30]. The data was mapped at each site according to public Radiation Oncology Ontology [31].

Statistics and model building were performed in Statistica version 13 (Dell Inc., Tulsa, OK) and MATLAB R2015b (The Mathworks Inc., Natick, MA). Significance was assumed for p values smaller than 0.05. TRIPOD reporting guidelines were followed [32].

Results

Patient and treatment characteristics of the 4 datasets are listed in Table 1. The development dataset had 1 and 2 year overall survival probabilities of 61.3% and 38.7%. Nonlinear transformations of continuous variables were not seen to significantly improve the likelihood of univariable associations. In univariable logistic regression for both the 12 and 24 month mortality endpoints and time-to-event Cox regression, the variables gross tumour volume (GTV), mean heart dose (MHD) and current smoker were significantly associated with increased mortality (Appendix A). Kaplan–Meier curves (Fig. 1) confirmed that GTV, with an association with mortality detectable as early as 2 months after treatment, and smoking, with an association detectable from 6 months after treatment, had a consistent impact on mortality risk up to 5 years. However, the survival curves show that the increased mortality

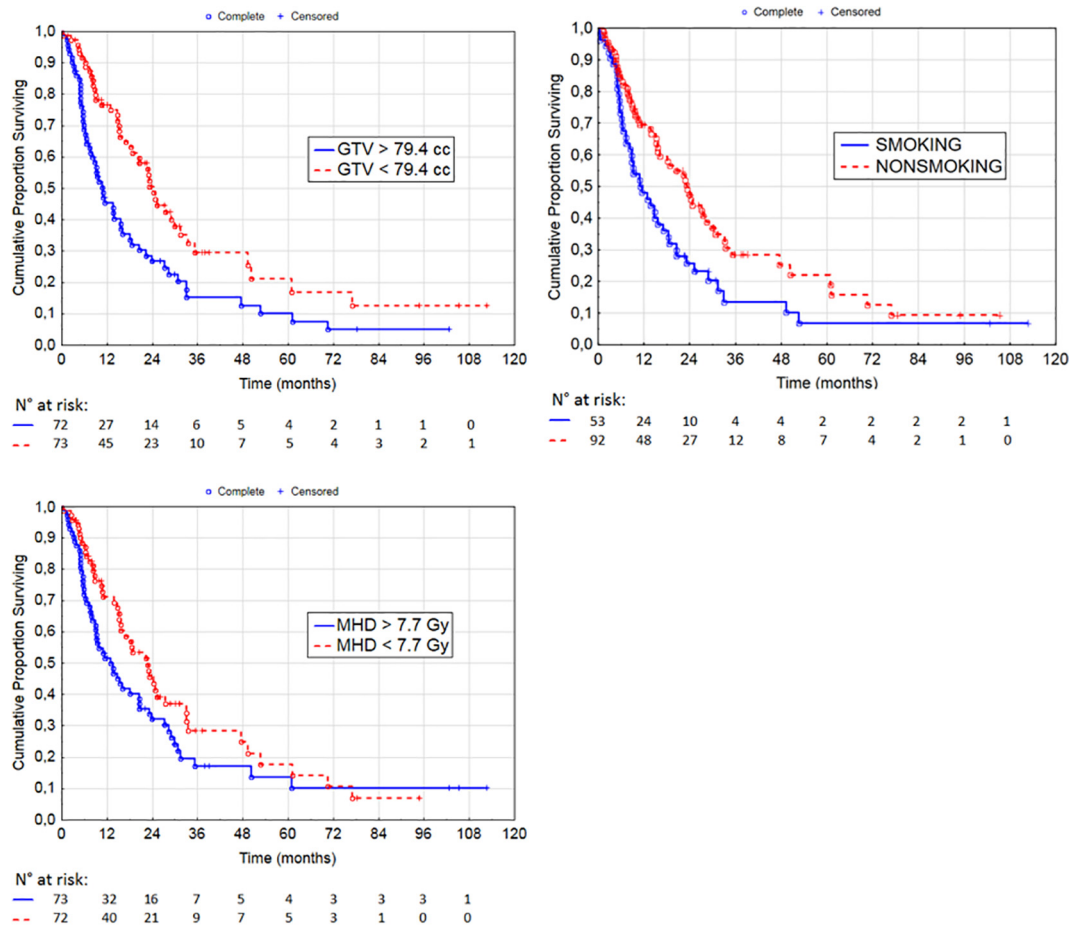**Fig. 1.** Kaplan–Meier curves for overall survival in the development set of 145 patients. Numbers at risk are indicated below the graphs and the follow-up time starts at end of RT. Groups based on gross tumour volume (GTV) (*upper left graph*), current smoking (*upper right graph*) and mean heart dose (MHD) (*lower left graph*). The cut-offs used for GTV and MHD were the median values for GTV of 79.4 cc and for MHD of 7.7 Gy.
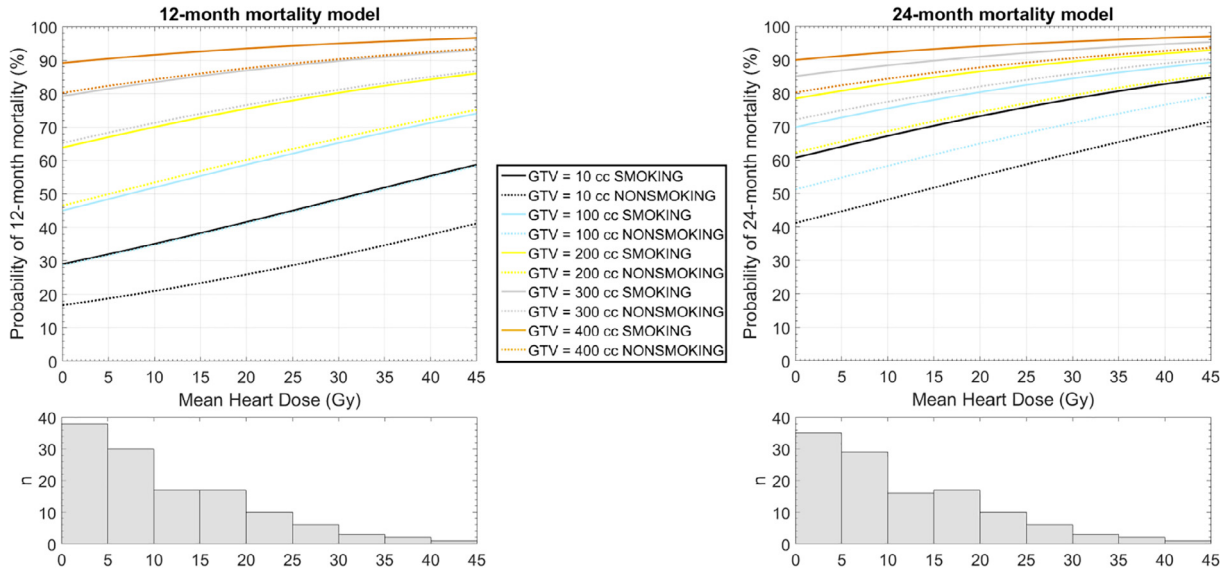
Fig. 2. GTV-smoker-MHD logistic regression models. Graphical representation of the GTV-smoker-MHD logistic regression models predicting 12 month (*left graph*) and 24 month (*right graph*) mortality risk in the development dataset of 145 patients. Mean Heart Dose (MHD) dependence of the risks for representative patient categories based on the covariates Gross Tumour Volume (GTV) and current smoking status. Predicted probabilities can be calculated using following formula: $P = (1 + e^{-S})^{-1}$ with $S = -1.68 + 0.0077*GTV + 0.71*SMOKER$ (Yes = 1/No = 0) + 0.0277*MHD for 12 month mortality risk and $S = -0.40 + 0.0045*GTV + 0.79*SMOKER$ (Yes = 1/No = −1) + 0.0283*MHD for 24 month mortality risk. The distribution of MHD values in the dataset (n = absolute number of patients) is shown at the bottom of the graphs.

Table 2

Multivariable logistic regression prediction models for 12 and 24 month mortality optimized in the development set. Predicted risks can be calculated as $P = (1 + e^{-S})^{-1}$ with $S = \beta_0 + \sum_{i=1}^{n} \beta_i x_i$. 4 models combining the most frequently selected covariates using 100 times repeated 5-fold cross-validation (Appendix B). Covariates (model coefficients, OR and p value), discriminative power (AUC and discrimination slope (average prediction difference between patients with and without the outcome)) and goodness of fit (Nagelkerke R^2). The GTV-smoker-MHD models are depicted in Fig. 2. Calibration of these models is depicted in Fig. 3 (24 month mortality) and Appendix E (12 month mortality).

	Model coefficient β		OR (95% CI)	p value	AUC (95% CI)	Discrim. slope	Nagelkerke R^2
	Original	Shrunk (*factor f*)					
12 month mortality prediction models							
GTV model		f = 1.15			0.74 (0.65; 0.82)	0.142	0.164
Intercept β₀	−1.229	−1.36			Optimism-corrected:		
GTV (+1 cc)	0.00891	0.0102	1.0090 (1.0036; 1.014)	0.0011	0.737 (0.737; 0.737)		
GTV-smoker model		f = 0.89			0.75 (0.66; 0.83)	0.172	0.205
Intercept β₀	−1.59	−1.45			Optimism-corrected:		
GTV (+1 cc)	0.00886	0.0079	1.0089 (1.0035; 1.014)	0.0012	0.745 (0.709; 0.758)		
Current smoker (yes vs no)	0.83	0.739	2.294 (1.056; 4.982)	0.036			
GTV-MHD model		f = 1.05			0.74 (0.65; 0.82)	0.162	0.189
Intercept β₀	−1.491	−1.54			Optimism-corrected:		
GTV (+1 cc)	0.00761	0.0080	1.0076 (1.0021; 1.013)	0.0070	0.733 (0.684; 0.747)		
MHD (+1 Gy)	0.0343	0.0359	1.035 (0.993; 1.079)	0.11			
GTV-smoker-MHD model		f = 0.92			0.77 (0.67; 0.84)	0.188	0.223
Intercept β₀	−1.795	−1.68			Optimism-corrected:		
GTV (+1 cc)	0.00775	0.0071	1.0078 (1.0022; 1.0134)	0.0061	0.745 (0.697; 0.763)		
Current smoker (yes vs no)	0.772	0.710	2.163 (0.987; 4.742)	0.054			
MHD (+1 Gy)	0.0301	0.0277	1.031 (0.988; 1.075)	0.16			
24 month mortality prediction models							
GTV model		f = 1.22			0.70 (0.60; 0.78)	0.068	0.080
Intercept β₀	0.0329	−0.088			Optimism-corrected:		
GTV (+1 cc)	0.00639	0.0078	1.0064 (1.0007; 1.012)	0.028	0.698 (0.698; 0.698)		
GTV-smoker model		f = 0.87			0.72 (0.63; 0.80)	0.110	0.139
Intercept β₀	−0.330	−0.21			Optimism-corrected:		
GTV (+1 cc)	0.00630	0.0055	1.0063 (1.00065; 1.012)	0.029	0.717 (0.704; 0.731)		
Current smoker (yes vs no)	0.979	0.852	2.663 (1.152; 6.157)	0.022			
GTV-MHD model		f = 0.97			0.69 (0.60; 0.77)	0.087	0.109
Intercept β₀	−0.248	−0.22			Optimism-corrected:		
GTV (+1 cc)	0.0051	0.0049	1.0051 (0.999; 1.011)	0.088	0.685 (0.589; 0.711)		
MHD (+1 Gy)	0.0378	0.0367	1.0385 (0.991; 1.088)	0.11			
GTV-smoker-MHD model		f = 0.87			0.72 (0.63; 0.80)	0.121	0.158
Intercept β₀	−0.546	−0.40			Optimism-corrected:		
GTV (+1 cc)	0.0052	0.0045	1.0052 (0.999; 1.011)	0.079	0.708 (0.639; 0.738)		
Current smoker (yes vs no)	0.912	0.793	2.490 (1.067; 5.809)	0.035			
MHD (+1 Gy)	0.0325	0.0283	1.0331 (0.985; 1.084)	0.19			

Abbreviations: GTV: Gross Tumour Volume; MHD: Mean Heart Dose; OR: Odds Ratio; CI: Confidence Interval; AUC: Area Under the Curve; Discrim. slope: Discrimination slope; R^2: Coefficient of determination.

Table 3

External validation of the prediction models of 12 and 24 month mortality. Metrics describing the discrimination (AUC and discrimination slope) and calibration (calibration-in-the-large and calibration slope, i.e., intercept and slope of the predicted versus observed mortality calibration curves) of the models in 3 external validation datasets. Calibration plots are depicted in Fig. 3 and Appendix E for the GTV-smoker-MHD models at 24 and 12 month, respectively. The appropriate model update was selected according to a closed testing procedure described by Vergouwe et al. [28]. It is reported whether the external datasets contain evidence to update the model intercept (recalibration-in-the-large), to additionally update the model slope (recalibration) or to re-estimate all model coefficients (revision). Updated calibration plots and regression formulae are shown in Appendix F for the GTV-smoker-MHD models.

| External validation sets | The Christie (878 patients) | | | | | Erasmus MC (127 patients) | | | | | UZ Leuven (96 patients) | | | | |
Prediction models	AUC	Discr. slope	Calibr.-in-the-large	Calibr. slope	Model update	AUC	Discr. slope	Calibr.-in-the-large	Calibr. slope	Model update	AUC	Discr. slope	Calibr.-in-the-large	Calibr. slope	Model update
12 month mortality															
GTV model	0.63	0.054	0.13	0.63	Revis.	0.51	−0.009	−1.47	−0.04	Slope	0.59	0.065	−1.39	0.24	Slope
GTV-smoker model	0.59	0.054	0.14	0.48	Revis.	NA	NA	NA	NA	NA	0.63	0.078	−1.40	0.37	Slope
GTV-MHD model	0.63	0.065	0.01	0.65	Slope	0.47	−0.025	−1.44	−0.12	Slope	0.61	0.079	−1.36	0.39	Slope
GTV-smoker-MHD model	0.61	0.070	0.02	0.54	Revis.	NA	NA	NA	NA	NA	0.65	0.094	−1.37	0.48	Interc.
24 month mortality															
GTV model	0.64	0.038	0.34	1.07	Interc	0.68	0.079	−0.73	0.63	Interc.	0.57	0.030	−1.02	0.21	Slope
GTV-smoker model	0.61	0.021	−0.20	0.54	Revis.	NA	NA	NA	NA	NA	0.60	0.051	−1.07	0.43	Interc.
GTV-MHD model	0.64	0.051	0.22	0.91	Interc	0.67	0.069	−0.75	0.88	Interc.	0.61	0.042	−1.00	0.53	Interc.
GTV-smoker-MHD model	0.61	0.055	0.34	0.65	Revis.	NA	NA	NA	NA	NA	0.64	0.064	−1.06	0.64	Interc.

Abbreviations: GTV: Gross Tumour Volume; MHD: Mean Heart Dose; AUC: Area Under the Curve; Discrim. slope: Discrimination slope; Calibr. Slope: Calibration slope. Interc.: Intercept; Revis.: Revision.

---

associated with higher MHD is most pronounced between 6 and 24 months after treatment.

In repeated cross-validation of model development (Appendix B), the GTV covariate was selected in 94% of logistic regression models, followed by current smoker (81%) and MHD (44%). MLD was selected in only 8% of the models. At both time points, the most frequently developed model was the GTV-smoker model (combining GTV and current smoker covariates) followed by the GTV-smoker-MHD model (Fig. 2). Further analysis therefore focused on these and on related (GTV-only and GTV-MHD) models, in order to study the additional benefit of each covariate. Final logistic regression model coefficients are reported in Table 2. Performance metrics showed only minimal differences in discrimination between the models (ΔAUC < 0.03). Optimism-corrected AUCs were 0.74 (95% CI: 0.66–0.78) and 0.69 (95% CI: 0.55–0.74) for modelling at 12 and 24 months, respectively. Cox regression models (Appendix C) showed even smaller c-statistic variations and worsened when adding the MHD covariate. However, the heterogeneity of the development dataset in terms of follow-up times in the different treatment periods might have biased the variable selection. The Cox models were therefore not studied further in the validation steps.

In the 3 external validation cohorts, survival probabilities varied between 62.0% and 81.2% at 1 year and between 35.1% and 54.4% at 2 years. The 24 month mortality prediction models mostly showed higher validation AUC than the 12 month mortality prediction models (Table 3). In all 3 validation datasets, 24 month mortality models including MHD proved to be robust as they required an update of the model intercept only, except for 1 model, while most other models required recalibration or revision in at least one of the datasets. The 24 month mortality GTV-smoker-MHD model validations performed best in terms of discrimination (0.61–0.64 before and 0.64–0.65 after appropriate model update, respectively), but the improvement over a GTV-only model was limited (ΔAUC = 0.01–0.07). This model showed a good calibration (slope 0.64–0.65, Fig. 3).

Discussion

To the best of our knowledge, this is the first study to build and externally validate prediction models of mortality at 1 and 2 years after radical chemoradiotherapy for NSCLC, taking into account clinical and dosimetric variables for more than 1000 patients from different European institutions. The most heterogeneous dataset in terms of prescribed radiation doses, tumour stages and radiotherapy techniques, resulting in large heart dose variability, was used as the development set. We observed a strong GTV dependence of mortality risk in logistic and Cox regression modelling, while current smoker and MHD had limited additional impact on the performance of the prediction model. Based on the analysis of 3 external validation datasets, the prediction model combining GTV, current smoker and MHD most robustly predicts mortality risk, especially at 24 months after the end of treatment (AUC = 0.61–0.64 with good calibration slopes). The 12 month mortality models did not translate well to external datasets (strongly degraded AUC and calibration metrics). This could partly be explained by the first year mortality exhibiting a lower dependence on GTV in validation datasets containing recent treatments (calibration slope of GTV-only model validation <0.3 in Erasmus MC and UZ Leuven datasets). From the studied models, the GTV-smoker-MHD model predicting mortality at 24 months after treatment could thus be recommended. However, the optimal endpoint may also depend on the preferred accuracy measure, with 12 month mortality prediction models having higher negative predictive values (Appendix D). An institution-specific validation focusing on an adjustment of the baseline mortality risk remains of paramount importance
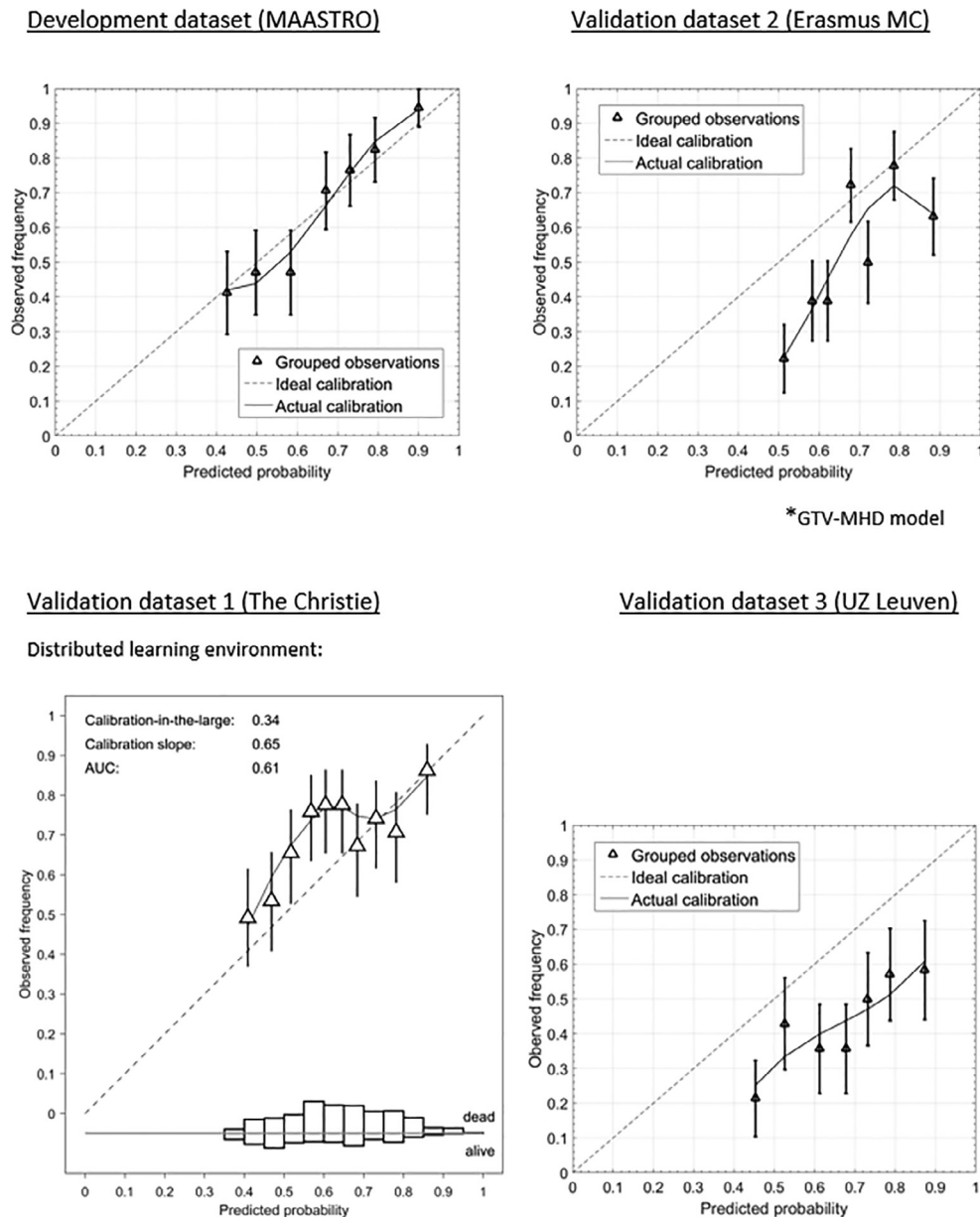
Fig. 3. Calibration plots for the 24 month mortality GTV-smoker-MHD prediction model. Calibration in the development set *(upper left)*, the Christie external validation cohort 1 *(lower left)*, Erasmus MC external validation cohort 2 (GTV-MHD model as no individual smoking information was present) *(upper right)* and UZ Leuven external validation cohort 3 *(lower right)*. In cohort 2 and 3, there was only evidence for updating the model intercept, e.g. to correct for the lower mortality frequency visible in the plots. Calibration plots after model updating are reported in Appendix F.

before any clinical application of the presented prediction models. Our Cox regression models might be suboptimal for the prediction of mortality in the first 2 years after treatment. The cross validation results of Appendix B show that using Cox regression other variables were often selected than using logistic regression, e.g. the Cox regression models including the WHO covariate would result in a worse performance for the prediction of 2 year mortality risks as WHO was practically never selected in the 2 year logistic regression model.

A major advantage of our prediction models is the highly relevant mortality endpoint and the selection of an optimal post-treatment time point, chosen to maximize the effect size of an actionable dosimetric variable (MHD). The Kaplan–Meier curves suggested the MHD to contain information associated with mortality risk between 6 and 24 months after treatment. An improvement in discrimination when adding MHD to the prediction

model was only observed in approximately one third of cross validation model developments, while it was small but consistently observed in two validation datasets. Including the MHD covariate was also important for the improvement of the calibration slopes observed in external validation. According to our data, only a limited subset of non-smoking patients with small tumour volumes associated with higher heart doses (i.e., located close to the heart) might have a detectable survival benefit from reductions of the MHD. For example, a MHD reduction from 20 Gy to 10 Gy would decrease the 24 month absolute mortality risk by 7.1% (12.8% in relative terms) for a 10 cc GTV in a non-smoking patient. A similar MHD reduction for a 200 cc GTV in a smoker would result in a substantially lower absolute survival gain of 3.7% (4.2% in relative terms).

Planning studies have shown that MHD reductions of at least 50% are commonly achievable with proton therapy [33–35]. Based

on these data, randomized trials could investigate the potential advantage of proton therapy with 24 month mortality as the primary endpoint in patient groups most at risk. The good calibration slope validation shown for our 24 month mortality prediction model is crucial in order to translate MHD reductions into improved survival on a population basis. However, improvement of the model accuracy is required to limit the number of patients who needlessly would be included in such study. Improvement could be achieved by the analysis of additional variables such as tumour location [3] and treatment technique [36] to strengthen our conclusions on MHD as an independent risk factor of mortality at 1 and 2 years after treatment. Radiomics analyses have shown promising image-based survival discrimination in lung cancer and should be part of future modelling strategies [37]. Finally, the models should be updated based on the inclusion of patients treated with the new standard of care of immune checkpoint inhibitor following concurrent chemoradiotherapy [38].

The GTV was shown in previous studies to be strongly associated with lung cancer survival [8,13,23,24,39]. Our models did not select WHO performance status [8,13,16,17], possibly because only fit patients were selected for chemoradiotherapy in the development dataset. In RTOG0617, a survival detriment was observed for patients with higher heart doses, from the first month after treatment, which is in line with our Kaplan–Meier curves. Our data suggest a strong association of smoking with mortality, which is higher than that of the heart dose. Although we could only obtain data on the smoking status at the time of initiation of radiotherapy, this finding supports the importance of smoking cessation programs in all lung cancer patients, as is reflected in guidelines [22]. Similarly, in a study estimating the risks of breast cancer radiotherapy, smoking was seen to dramatically increase the MHD-related risk of cardiac mortality [40].

This study has some limitations. Our analysis did not allow to draw conclusions on the causal relation between cardiac dose and mortality. Ideally, our prediction model should be evaluated in a randomized trial. The development dataset was heterogeneous in terms of time period of treatment, which was associated to some changes in the standard treatment. While this might not be the ideal situation for model building, it had the advantage of resulting in generalizable 24 month mortality model coefficients, with acceptable performance in all external validation datasets. Even though a large proportion of patients received radiotherapy-only treatments in the Christie cohort, this did not negatively influence model discrimination and calibration. Another caveat is that the Christie dataset had survival times calculated from the start instead of the end of treatment. This 1–1.5 month bias in the assessment of mortality is not likely to have influenced model discrimination analysis as the bias was approximately the same in all patients. Furthermore, a potential bias from the inclusion of cases with complete data into the development dataset could not be excluded. The available follow-up in our development dataset was a major limitation for the Cox model fits. The most recent part of the development dataset (2014–2016) had a median follow-up time of 22.6 months, while this was 102.6 months for the older part (2003–2006). The Cox models reported in Appendix C might thus contain risk factors that are biased towards the older part of the dataset.

Initial treatment plan dosimetric data was used, while it has been shown that the interfractional average MHD variation is 1.2 Gy [41]. This should nevertheless not change the conclusions. Only physical MLD and MHD doses were available. With almost all treatments delivered in at least 20 fractions and an upper MHD value of 46.0 Gy, the impact of a recalculation in 2 Gy equivalent doses is expected to be limited. A complete heart DVH analysis might have enhanced dose response modelling, but no previous studies have shown that the selection of one specific heart dose-volume metric significantly improves associations with outcome when compared to another heart dose-volume metric [6,10,11]. Finally, a refinement of our understanding of individual heart substructure radiosensitivities should come from prospective studies collecting cardiac imaging and circulating biomarkers.

In conclusion, we developed an externally validated, moderately discriminating and well-calibrated prediction model of 24 month mortality after radical chemoradiotherapy in NSCLC patients. The model shows that patient prognosis is strongly related to tumour volume and reveals an important association with smoking. It allows the identification of individual patients (i.e., those with small central tumours) for whom a reduction of the heart dose might be beneficial.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2019.09.005.

References

[1] Auperin A, Le Pechoux C, Rolland E, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. J Clin Oncol 2010;28:2181–90.

[2] Bradley JD, Paulus R, Komaki R, 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.

[3] Bradley JD, Hu C, Komaki RU, et al. Long-term results of RTOG 0617: a randomized phase 3 comparison of standard dose versus high dose conformal chemoradiation therapy +/- cetuximab for stage III NSCLC. Int J Rad Oncol Biol Phys 2017;99:S105.

[4] Eberhardt WEE, Pöttgen C, Gauler TC, et al. Phase III study of surgery versus definitive concurrent chemoradiotherapy boost in patients with resectable stage IIIA(N2) and selected IIIB non-small-cell lung cancer after induction chemotherapy and concurrent chemoradiotherapy (ESPATUE). J Clin Oncol 2015;33:4194–201.

[5] Senan S, Brade A, Wang LH, et al. PROCLAIM: randomized phase III trial of pemetrexed-cisplatin or etoposide-cisplatin plus thoracic radiation therapy followed by consolidation chemotherapy in locally advanced nonsquamous non-small-cell lung cancer. J Clin Oncol 2016;34:953–62.

[6] Speirs CK, DeWees TA, Rehman S, et al. Heart dose is an independent dosimetric predictor of overall survival in locally advanced non-small cell lung cancer. J Thorac Oncol 2017;12:293–301.

[7] De Ruysscher D, Faivre-Finn C, Moeller D, et al. European Organization for Research and Treatment of Cancer (EORTC) recommendations for planning and delivery of high-dose, high precision radiotherapy for lung cancer. Radiother Oncol 2017;124:1–10.

[8] Tucker SL, Liu A, Gomez D, 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.

[9] Guberina M, Eberhardt W, Stuschke M, et al. Heart dose exposure as prognostic marker after radiotherapy for resectable stage IIIA/B non-small-cell lung cancer: secondary analysis of a randomized trial. Ann Oncol 2017;28:1084–9.

[10] Dess RT, Sun Y, Matuszak MM, 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.

[11] Wang K, Eblan MJ, Deal AM, 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.

[12] Stam B, Peulen H, Guckenberger M, et al. Dose to heart substructures is associated with non-cancer death after SBRT in stage I-II NSCLC patients. Radiother Oncol 2017;123:370–5.

[13] McWilliam A, Kennedy J, Hodgson C, Vasquez Osorio 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.

[14] Johnson CN, Price GJ, Faivre-Finn C, Aznar MC, van Herk M. Residual setup errors towards the heart after image guidance linked with poorer survival in lung cancer patients: do we need stricter IGRT protocols?. Int J Radiat Oncol Biol Phys 2018.

[15] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.

[16] Jochems A, Deist TM, El Naqa I, et al. Developing and validating a survival prediction model for NSCLC patients through distributed learning across 3 countries. Int J Radiat Oncol Biol Phys 2017;99:344–52.

[17] Dehing-Oberije C, De Ruysscher D, van der Weide H, et al. Tumor volume combined with number of positive lymph node stations is a more important prognostic factor than TNM stage for survival of non-small-cell lung cancer patients treated with (chemo)radiotherapy. Int J Radiat Oncol Biol Phys 2008;70:1039–44.

[18] van Elmpt W, De Ruysscher D, van der Salm A, et al. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. Radiother Oncol 2012;104:67–71.

[19] Wolthaus JW, Schneider C, Sonke JJ, et al. Mid-ventilation CT scan construction from four-dimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. Int J Radiat Oncol Biol Phys 2006;65:1560–71.

[20] Verbakel WF, van Reij E, Ladenius-Lischer I, Cuijpers JP, Slotman BJ, Senan S. Clinical application of a novel hybrid intensity-modulated radiotherapy technique for stage III lung cancer and dosimetric comparison with four other techniques. Int J Radiat Oncol Biol Phys 2012;83:e297–303.

[21] Johnson C, Price G, Khalifa J, et al. A method to combine target volume data from 3D and 4D planned thoracic radiotherapy patient cohorts for machine learning applications. Radiother Oncol 2018;126:355–61.

[22] Postmus PE, Kerr KM, Oudkerk M, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2017;28:iv1-iv21.

[23] Bradley JD, Ieumwananonthachai N, Purdy J, et al. Gross tumor volume, critical prognostic factor in patients treated with three-dimensional conformal radiatio therapy for non-small cell lung carcinoma. Int J Rad Oncol Biol Phys 2002;52:49–57.

[24] Werner-Wasik M, Swann RS, Bradley J, et al. Increasing tumor volume is predictive of poor overall and progression-free survival: secondary analysis of the Radiation Therapy Oncology Group 93–11 phase I-II radiation dose-escalation study in patients with inoperable non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2008;70:385–90.

[25] Zhang Z. Variable selection with stepwise and best subset approaches. Ann Transl Med 2016;4:136.

[26] Harrell F. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. Springer; 2001.

[27] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38.

[28] Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 2017;36:4529–39.

[29] Jochems A, Deist TM, van Soest J, et al. Distributed learning: developing a predictive model based on data from multiple hospitals without data leaving the hospital – a real life proof of concept. Radiother Oncol 2016;121:459–67.

[30] Damiani A, Vallati M, Gatta R, et al. Distributed learning to protect privacy in multi-centric clinical studies. Artif Intell Med 2015:65–75.

[31] BioPortal. Radiation Oncology Ontology. 2018. https://bioportal.bioontology.org/ontologies/ROO2018.

[32] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13:1.

[33] Zhang X, Li Y, Pan X, et al. Intensity-modulated proton therapy reduces the dose to normal tissue compared with intensity-modulated radiation therapy or passive scattering proton therapy and enables individualized radical radiotherapy for extensive stage IIIB non-small-cell lung cancer: a virtual clinical study. Int J Radiat Oncol Biol Phys 2010;77:357–66.

[34] Chang JY, Li H, Zhu XR, et al. Clinical implementation of intensity modulated proton therapy for thoracic malignancies. Int J Radiat Oncol Biol Phys 2014;90:809–18.

[35] Chang JY, Zhang W, Komaki R, et al. Long-term outcome of phase I/II prospective study of dose-escalated proton therapy for early-stage non-small cell lung cancer. Radiother Oncol 2017;122:274–80.

[36] Chun SG, Hu C, Choy H, 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;35:56–62.

[37] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

[38] Antonia SJ, Villegas A, Daniel D, et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. N Engl J Med 2018.

[39] van Baardwijk A, Wanders S, Boersma L, et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. J Clin Oncol 2010;28:1380–6.

[40] Taylor C, Correa C, Duane FK, et al. Estimating the risks of breast cancer radiotherapy: evidence from modern radiation doses to the lungs and heart and from previous randomized trials. J Clin Oncol 2017;35:1641–9.

[41] Jan N, Guy C, Reshko LB, Hugo GD, Weiss E. Lung and heart dose variability during radiation therapy of non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2017;98:683–90.