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

Citation for published version (APA):

Document status and date:
Published: 01/11/2020

DOI:
10.1016/j.radonc.2019.09.005

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:
repository@maastrichtuniversity.nl
providing details and we will investigate your claim.

Download date: 17 Sep. 2023
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 Arredouania a, Maarten Lambrechts a, Joost Nuyttens f, Corinne Faiivre-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 d Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

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 Arredouania a, Maarten Lambrechts a, Joost Nuyttens f, Corinne Faiivre-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 d Department of Radiotherapy, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

A R T I C L E   I N F O

Article history:
Received 15 May 2019
Received in revised form 4 September 2019
Accepted 4 September 2019
Available online 20 September 2019

Keywords:
Mean heart dose
Overall survival
Prediction model
Lung cancer
Proton therapy

A B S T R A C T

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
guidance datasets have identified a region at the base (i.e., superi-
osterior 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 includ-
ing the impact of heart dose would allow personalization of plan-
ning dose constraints. This could facilitate the selection of high-
risk patients that may benefit from emerging radiotherapy tech-
nologies, enabling high precision delivery of the radiation dose
(i.e., lowering the heart dose) such as proton therapy and the mag-
netic resonance-guided linear accelerator radiotherapy (MR-linac)
[15]. Published externally validated prediction models of lung can-
cer 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
electronically with area under the curve around 0.60. Others studied
the influence of the heart dose on mortality in a monocentric data-
set, but did not include important risk factors such as tumour vol-
ume [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 realis-
tic 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 regres-
sion mortality prediction models at 1 and 2 year after treat-
ment in radically treated NSCLC cohorts from 4 separate institu-
tions. 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 chemoradio-
therapy at MAASTRO Clinic during two time periods (previously
collected datasets from 2003–2006 and 2014–2016 periods) con-
stituted 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 hun-
dred forty five patients with complete data were included. Pre-
scription 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 isotropically dose-escalated radio-
therapy up to 106.4 Gy (24 fractions, positron emission tomogra-
phy (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
(ITT) 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 radio-
therapy (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 radiother-
apy 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 cen-
ters have approved the study.

Clinical and dosimetric variables were chosen based on their
importance as prognostic factors for overall survival in the litera-
ture: baseline World Health Organization performance status
[8,13,16,17], age at start of treatment [16], current smoker at time
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 prognos-
tic 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. Sur-
vival 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 mortal-
ity. 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 trans-
formations (log, square root, inverse transformations, etc.) of con-
tinuous variables were tested in the univariable analyses.
Backward stepwise model building processes (including all vari-
ables) 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 recei-
ver operating characteristic curve (AUC) or c-statistic. Model cali-
bration 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.

<table>
<thead>
<tr>
<th></th>
<th>Development set</th>
<th>External validation sets</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>65 (39–88)</td>
<td>71 (32–93)</td>
<td>62 (30–80)</td>
</tr>
<tr>
<td>Tumour T stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13 (9.0 %)</td>
<td>102 (11.6 %)</td>
<td>19 (15.0 %)</td>
</tr>
<tr>
<td>2</td>
<td>42 (29.0 %)</td>
<td>290 (33.0 %)</td>
<td>20 (15.8 %)</td>
</tr>
<tr>
<td>3</td>
<td>14 (9.7 %)</td>
<td>241 (27.5 %)</td>
<td>25 (19.7 %)</td>
</tr>
<tr>
<td>4</td>
<td>58 (40.0 %)</td>
<td>204 (23.2 %)</td>
<td>59 (46.5 %)</td>
</tr>
<tr>
<td>Unknown</td>
<td>18 (12.4 %)</td>
<td>41 (4.7 %)</td>
<td>4 (3.2 %)</td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>31 (21.4 %)</td>
<td>272 (31.0 %)</td>
<td>20 (15.8 %)</td>
</tr>
<tr>
<td>1</td>
<td>2 (1.4 %)</td>
<td>114 (13.0 %)</td>
<td>10 (7.9 %)</td>
</tr>
<tr>
<td>2</td>
<td>55 (37.9 %)</td>
<td>298 (33.9 %)</td>
<td>79 (62.2 %)</td>
</tr>
<tr>
<td>3</td>
<td>40 (27.6 %)</td>
<td>165 (18.8 %)</td>
<td>18 (14.2 %)</td>
</tr>
<tr>
<td>Unknown</td>
<td>17 (11.7 %)</td>
<td>29 (3.3 %)</td>
<td>NA</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>50 (34.5 %)</td>
<td>418 (47.6 %)</td>
<td>46 (36.2 %)</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>31 (21.4 %)</td>
<td>223 (25.4 %)</td>
<td>49 (38.6 %)</td>
</tr>
<tr>
<td>NSCLC NOS</td>
<td>16 (11.0 %)</td>
<td>94 (10.7 %)</td>
<td>NA</td>
</tr>
<tr>
<td>Large cell</td>
<td>37 (25.5 %)</td>
<td>7 (0.8 %)</td>
<td>32 (25.2 %)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (2.1 %)</td>
<td>31 (3.5 %)</td>
<td>NA</td>
</tr>
<tr>
<td>Unknown</td>
<td>8 (5.5 %)</td>
<td>105 (12.0 %)</td>
<td>NA</td>
</tr>
<tr>
<td>WHO performance status at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>53 (36.6 %)</td>
<td>228 (26.0 %)</td>
<td>118 (92.9 %)</td>
</tr>
<tr>
<td>Stopped/never smoker</td>
<td>92 (63.5 %)</td>
<td>418 (47.6 %)</td>
<td>6 (4.7 %)**</td>
</tr>
<tr>
<td>Unknown</td>
<td>232 (26.4 %)</td>
<td>NA</td>
<td>3 (2.4 %)</td>
</tr>
<tr>
<td>Smoking status at diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>53 (36.6 %)</td>
<td>228 (26.0 %)</td>
<td>118 (92.9 %)</td>
</tr>
<tr>
<td>Stopped/never smoker</td>
<td>92 (63.5 %)</td>
<td>418 (47.6 %)</td>
<td>6 (4.7 %)**</td>
</tr>
<tr>
<td>Unknown</td>
<td>232 (26.4 %)</td>
<td>NA</td>
<td>3 (2.4 %)</td>
</tr>
<tr>
<td>Chemotherapy treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concurrent</td>
<td>59 (40.7 %)</td>
<td>182 (20.7 %)</td>
<td>111 (87.4 %)</td>
</tr>
<tr>
<td>Sequential</td>
<td>86 (59.3 %)</td>
<td>143 (16.3 %)</td>
<td>14 (11.0 %)</td>
</tr>
<tr>
<td>RT alone or unknown</td>
<td>553 (63.0 %)</td>
<td>2 (1.6 %)</td>
<td>2 (1.5 %)</td>
</tr>
<tr>
<td>Treatment technique</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D-CRT</td>
<td>90 (62.1 %)</td>
<td>280 (31.9 %)</td>
<td>127 (100.0 %)</td>
</tr>
<tr>
<td>IMRT</td>
<td>NA</td>
<td>583 (66.4 %)</td>
<td>NA</td>
</tr>
<tr>
<td>VMAT</td>
<td>45 (31.0 %)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Hybrid VMAT</td>
<td>10 (6.9 %)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Unknown</td>
<td>NA</td>
<td>15 (1.7 %)</td>
<td>NA</td>
</tr>
<tr>
<td>GTV volume (cc)</td>
<td>79.4 (0.3–996.8)</td>
<td>45.5 (0.3–501.5)</td>
<td>84.9 (0.5–510.2)</td>
</tr>
<tr>
<td>Dose per fraction (Gy)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>26 (17.9 %)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1.8</td>
<td>37 (25.5 %)</td>
<td>NA</td>
<td>1 (0.8 %)</td>
</tr>
<tr>
<td>2</td>
<td>69 (47.6 %)</td>
<td>182 (20.7 %)</td>
<td>115 (90.6 %)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>13 (9.0 %)</td>
<td>696 (79.3 %)</td>
<td>286 (23.4 %)</td>
</tr>
<tr>
<td>Unknown</td>
<td>NA</td>
<td>NA</td>
<td>1 (1.0 %)</td>
</tr>
<tr>
<td>Mean Lung Dose (Gy physical dose)</td>
<td>15.6 (2.7–23.6)</td>
<td>13.8 (3.3–25.7)</td>
<td>15.5 (2.3–22.2)</td>
</tr>
<tr>
<td>Mean Heart Dose (Gy physical dose)</td>
<td>7.7 (0.1–45.2)</td>
<td>13.1 (0.0–35.4)</td>
<td>10.9 (0.0–46.0)</td>
</tr>
<tr>
<td>1 year survival probability (%)</td>
<td>61.3</td>
<td>62.0</td>
<td>78.7</td>
</tr>
<tr>
<td>&lt;2009</td>
<td>52.8</td>
<td>NA***</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;2009</td>
<td>76.9</td>
<td>NA</td>
<td>78.7</td>
</tr>
<tr>
<td>2 year survival probability (%)</td>
<td>38.7</td>
<td>35.1</td>
<td>48.0</td>
</tr>
<tr>
<td>&lt;2009</td>
<td>29.2</td>
<td>NA***</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;2009</td>
<td>61.5</td>
<td>35.1</td>
<td>48.0</td>
</tr>
<tr>
<td>Median survival time (month)</td>
<td>18.1</td>
<td>16.5</td>
<td>22.6</td>
</tr>
<tr>
<td>&lt;2009</td>
<td>13.9</td>
<td>NA***</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;2009</td>
<td>25.2</td>
<td>16.5</td>
<td>22.6</td>
</tr>
<tr>
<td>Median follow-up time (month)</td>
<td>35.5</td>
<td>39.1</td>
<td>67.0</td>
</tr>
<tr>
<td>&lt;2009</td>
<td>102.6</td>
<td>NA***</td>
<td>NA</td>
</tr>
<tr>
<td>&gt;2009</td>
<td>22.6</td>
<td>39.1</td>
<td>67.0</td>
</tr>
</tbody>
</table>


*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. 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 risk due to smoking was detectable as early as 6 months after treatment and the increased mortality risk due to GTV was detectable as early as 2 months after treatment.

![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.](image_url)
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: $P = \frac{1 + e^{-\beta_0 + \sum \beta_i x_i}}{C_0}$ with $S = C_0^{1.68 + 0.0077 \times \text{GTV} + 0.0045 \times \text{SMOKER \ (Yes = 1/No = 0)} + 0.0277 \times \text{MHD}}$ for 12 month mortality risk and $S = C_0^{0.40 + 0.0045 \times \text{GTV} + 0.79 \times \text{SMOKER \ (Yes = 1/No = 0)} + 0.0283 \times \text{MHD}}$ 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 $P = \frac{1 + e^{-\beta_0 + \sum \beta_i x_i}}{C_0}$ with $S = C_0^{\beta_0 + \sum \beta_i x_i}$. 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).

<table>
<thead>
<tr>
<th>Model</th>
<th>Model coefficient $\beta$</th>
<th>OR (95% CI)</th>
<th>$p$ value</th>
<th>AUC (95% CI)</th>
<th>Discrim. slope</th>
<th>Nagelkerke $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>12 month mortality prediction models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV model</td>
<td>$f = 1.15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-1.229$</td>
<td>$0.00891$</td>
<td>$0.0090$</td>
<td>$1.0036$; $1.014$</td>
<td>$0.0011$</td>
<td>$0.74$; $0.82$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$0.00102$</td>
<td>$1.00090$</td>
<td>$1.0036$; $1.014$</td>
<td>$0.0011$</td>
<td>$0.737$; $0.737$</td>
<td>Optimism-corrected:</td>
</tr>
<tr>
<td>GTV-Smoker model</td>
<td>$f = 0.89$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-1.59$</td>
<td>$0.00886$</td>
<td>$0.0079$</td>
<td>$1.0089$</td>
<td>$1.0035$; $1.014$</td>
<td>$0.0012$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-1.45$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker (yes vs no)</td>
<td>$0.83$</td>
<td>$0.739$</td>
<td>$2.294$</td>
<td>$1.056$; $4.982$</td>
<td>$0.036$</td>
<td>$0.74$; $0.82$</td>
</tr>
<tr>
<td>GTV-MHD model</td>
<td>$f = 1.05$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-1.491$</td>
<td>$0.00761$</td>
<td>$0.0080$</td>
<td>$1.0076$</td>
<td>$1.0021$; $1.013$</td>
<td>$0.0070$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-1.54$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MHD (+1 Gy)</td>
<td>$0.0343$</td>
<td>$1.035$</td>
<td>$1.0050$; $1.079$</td>
<td>$0.11$</td>
<td>$0.773$; $0.84$</td>
<td>Optimism-corrected:</td>
</tr>
<tr>
<td>GTV-Smoker-MHD model</td>
<td>$f = 0.92$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-1.795$</td>
<td>$0.00775$</td>
<td>$0.0071$</td>
<td>$1.0078$</td>
<td>$1.0022$; $1.013$</td>
<td>$0.0061$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-1.68$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker (yes vs no)</td>
<td>$0.772$</td>
<td>$0.710$</td>
<td>$2.163$</td>
<td>$0.987$; $4.742$</td>
<td>$0.054$</td>
<td>Optimism-corrected:</td>
</tr>
<tr>
<td>MHD (+1 Gy)</td>
<td>$0.0301$</td>
<td>$1.031$</td>
<td>$1.0050$; $1.075$</td>
<td>$0.16$</td>
<td>$0.777$; $0.84$</td>
<td>$0.188$</td>
</tr>
<tr>
<td><strong>24 month mortality prediction models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV model</td>
<td>$f = 1.22$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-0.329$</td>
<td>$0.00639$</td>
<td>$0.0078$</td>
<td>$1.0064$</td>
<td>$1.0007$; $1.012$</td>
<td>$0.028$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-0.21$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV-Smoker model</td>
<td>$f = 0.87$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-0.330$</td>
<td>$0.00630$</td>
<td>$0.0055$</td>
<td>$1.0063$</td>
<td>$1.00065$; $1.012$</td>
<td>$0.029$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker (yes vs no)</td>
<td>$0.979$</td>
<td>$0.852$</td>
<td>$2.663$</td>
<td>$1.152$; $6.157$</td>
<td>$0.022$</td>
<td>Optimism-corrected:</td>
</tr>
<tr>
<td>GTV-MHD model</td>
<td>$f = 0.97$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-0.248$</td>
<td>$0.0051$</td>
<td>$0.0049$</td>
<td>$1.0051$</td>
<td>$0.999$; $1.011$</td>
<td>$0.088$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-0.22$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MHD (+1 Gy)</td>
<td>$0.0378$</td>
<td>$1.0385$</td>
<td>$0.991$; $1.088$</td>
<td>$0.11$</td>
<td>Optimism-corrected:</td>
<td></td>
</tr>
<tr>
<td>GTV-Smoker-MHD model</td>
<td>$f = 0.87$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>$-0.546$</td>
<td>$0.0052$</td>
<td>$0.0045$</td>
<td>$1.0052$</td>
<td>$0.999$; $1.011$</td>
<td>$0.079$</td>
</tr>
<tr>
<td>GTV (+1 cc)</td>
<td>$-0.40$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker (yes vs no)</td>
<td>$0.912$</td>
<td>$0.793$</td>
<td>$2.490$</td>
<td>$1.067$; $5.809$</td>
<td>$0.035$</td>
<td>Optimism-corrected:</td>
</tr>
<tr>
<td>MHD (+1 Gy)</td>
<td>$0.0325$</td>
<td>$1.0331$</td>
<td>$0.985$; $1.084$</td>
<td>$0.19$</td>
<td>$0.708$; $0.738$</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>12 month mortality</th>
<th>24 month mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prediction models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV-smoker model</td>
<td>0.61</td>
<td>0.64</td>
</tr>
<tr>
<td>GTV-MHD model</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>GTV-smoker-MHD model</td>
<td>0.61</td>
<td>0.64</td>
</tr>
</tbody>
</table>

**AUC**: Area Under the Curve; **Discrimination**: Discrimination slope; **Calibr.**: Calibration slope; **Revis.**: Revision.

**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 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).
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 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

This work was supported by the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement 601826 (REQUITE). GJP and CFF gratefully acknowledge the support of Cancer Research UK via funding to the Cancer Research Manchester Centre [C147/ A18083 and A25254] and the support of the NIHR Biomedical Research Centre.

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


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


