International Journal of Radiation Oncology biology • physics

www.redjournal.org

Physics Contribution

Correlating Dose Variables with Local Tumor Control in Stereotactic Body Radiation Therapy for Early-Stage Non-Small Cell Lung Cancer: A Modeling Study on 1500 Individual Treatments

Rainer J. Klement, PhD,* Jan-Jakob Sonke, PhD,† Michael Allgäuer, MD,‡ Nicolaus Andratschke, MD,§ Steffen Appold, MD,‖ José Belderbos, MD, PhD,† Claus Belka, MD,¶ Oliver Blanck, PhD,# Karin Dieckmann, MD,** Hans T. Eich, MD,†† Frederick Mantel, MD,‡‡ Michael Eble, MD,§§ Andrew Hope, MD,‖‖ Anca L. Grosu, MD,¶¶ Meinhard Nevinny-Stickel, MD,## Sabine Semrau, MD,*** Reinhart A. Sweeney, MD,* Juliane Hörner-Rieber, MD,††† Maria Werner-Wasik, MD,§§§ Rita Engenhart-Cabillic, MD,‖‖‖ Hong Ye, PhD,¶¶¶ Inga Grills, MD,¶¶¶ and Matthias Guckenberger, MD§

**Department of Radiotherapy and Radiation Oncology, Leopoldina Hospital Schweinfurt, Schweinfurt, Germany; †Department of Radiation Oncology, Netherlands Cancer Institute, Amsterdam, Netherlands; ‡Department of Radiotherapy, Barmherzige Brüder Regensburg, Regensburg, Germany; §Department of Radiation Oncology, University Hospital Zurich, Zurich, Switzerland; ‖Department of Radiation Oncology, Technische Universität Dresden, Dresden, Germany; ¶Department of Radiation Oncology, University Hospital of Ludwig-Maximilians-University Munich, Munich, Germany; #Department of Radiation Oncology, Universitätsklinikum Schleswig-Holstein, Kiel, Germany; **Department of Radiotherapy, Medical University of Vienna, Vienna, Austria; ††Department of Radiotherapy, University Hospital Münster, Münster, Germany; ‡‡Department of Radiotherapy and Radiation Oncology, University Hospital Wuerzburg, Wuerzberg, Germany; §§Department of Radiation Oncology, RWTH Aachen University, Aachen, Germany; ‖‖Department of Radiation Oncology, University of Toronto and Princess Margaret Cancer Center, Toronto, Canada; ¶¶Department of Radiation Oncology, University Hospital Freiburg, Freiburg, Germany; ##Department of Radiation Oncology, Medical University Innsbruck, Innsbruck, Austria; ***Department of Radiation Oncology, Friedrich-Alexander University Erlangen-Nuremberg,*

Corresponding author: Rainer J. Klement, PhD; E-mail: rainer_klement@gmx.de

Inga Grills and Matthias Guckenberger made equal contributions to this study.

This research was partially supported through an Elekta research grant with 5 institutions being members of the Elekta Lung Research Group.

Disclosures: I.G. reports grants from Elekta during the conduct of the study. A.H. reports grants and nonfinancial support from Elekta during the conduct of the study. J.-J.S. has received research grants and royalties from Elekta Oncology, Ltd, and holds patents with royalties paid to Elekta.

Data sharing statement: Research data are not available at this time.

Supplementary material for this article can be found at https://doi.org/10.1016/j.ijrobp.2020.03.005.

Int J Radiation Oncol Biol Phys, Vol. 107, No. 3, pp. 579–586, 2020
0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2020.03.005

Erlangen, Germany; [†††]*Department of Radiation Oncology, University Hospital Heidelberg, Heidelberg, Germany;* [§§§]*Department of Radiation Oncology, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania;* [‖‖‖]*Department of Radiotherapy and Radiation Oncology, Phillips-University Marburg, Marburg, Germany; and* [¶¶¶]*Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, Michigan*

Received Nov 26, 2019. Accepted for publication Mar 2, 2020.

Background: Large variation regarding prescription and dose inhomogeneity exists in stereotactic body radiation therapy (SBRT) for early-stage non-small cell lung cancer. The aim of this modeling study was to identify which dose metric correlates best with local tumor control probability to make recommendations regarding SBRT prescription.

Methods and Materials: We combined 2 retrospective databases of patients with non-small cell lung cancer, yielding 1500 SBRT treatments for analysis. Three dose parameters were converted to biologically effective doses (BEDs): (1) the (near-minimum) dose prescribed to the planning target volume (PTV) periphery (yielding BED_{min}); (2) the (near-maximum) dose absorbed by 1% of the PTV (yielding BED_{max}); and (3) the average between near-minimum and near-maximum doses (yielding BED_{ave}). These BED parameters were then correlated to the risk of local recurrence through Cox regression. Furthermore, BED-based prediction of local recurrence was attempted by logistic regression and fast and frugal trees. Models were compared using the Akaike information criterion.

Results: There were 1500 treatments in 1434 patients; 117 tumors recurred locally. Actuarial local control rates at 12 and 36 months were 96.8% (95% confidence interval, 95.8%-97.8%) and 89.0% (87.0%-91.1%), respectively. In univariable Cox regression, BED_{ave} was the best predictor of risk of local recurrence, and a model based on BED_{min} had substantially less evidential support. In univariable logistic regression, the model based on BED_{ave} also performed best. Multivariable classification using fast and frugal trees revealed BED_{max} to be the most important predictor, followed by BED_{ave}.

Conclusions: BED_{ave} was generally better correlated with tumor control probability than either BED_{max} or BED_{min}. Because the average between near-minimum and near-maximum doses was highly correlated to the mean gross tumor volume dose, the latter may be used as a prescription target. More emphasis could be placed on achieving sufficiently high mean doses within the gross tumor volume rather than the PTV covering dose, a concept needing further validation.

Introduction

Stereotactic body radiation therapy (SBRT) is defined as a method of external beam radiation therapy in which a clearly defined extracranial target volume is treated with high precision and accuracy with a high radiation dose in a single or a few fractions with locally curative intent.[1] SBRT is increasingly used for the treatment of primary non-small cell lung cancer (NSCLC),[1] prostate cancer,[2] hepatic cancers,[3] and other primary malignancies as well as extracranial oligometastases.[4-6] The technological, quality, and dosimetry requirements for performing SBRT have been summarized in multiple national and international guidelines[1,7-9] and recently in detail in report 91 of the International Commission on Radiation Units and Measurements (ICRU), published in 2017.[10] According to the ICRU report 91, doses in SBRT should be prescribed to the planning target volume (PTV) covering the isodose surface. However, current SBRT practice frequently uses inhomogeneous dose distributions within the target volume, and large interinstitutional and intrainstitutional variation in the degree of dose inhomogeneity has been reported in planning studies[11] and in clinical routine[12-14]; this important issue of dose prescription remains unspecified in ICRU report 91. According to a recent review of ICRU report 91, "[t]his comes as no surprise as there is a lack of consensus on the most relevant dosimetric parameters influencing local tumor control."[15]

Many previous studies have either correlated the isocenter/near-maximum dose or the dose prescribed to the PTV periphery with tumor control probability (TCP).[16-20] However, these single dose-volume histogram parameters do not comprehensively describe the dose distribution within the PTV and are therefore insufficient in particular for reproducible inverse treatment planning. The aim of this paper is to identify whether a combination of both dose parameters would be better associated with local tumor control, which would allow better standardization of SBRT practice. This study is based on 2 large, retrospective patient databases that contain sufficiently large variation in SBRT practice to identify optimal SBRT planning characteristics.

Methods and Materials

Patients and treatments

This modeling study is based on 2 large retrospective databases of patients with early-stage NSCLC treated with

SBRT (1-10 fractions, median 3 fractions). One is the Elekta Collaborative Lung Research Group (ECLRG) database[16,21]; the other is the database of the working group "Stereotactic Radiotherapy and Radiosurgery" of the German Society for Radiation Oncology (DEGRO).[13,17] Except for 59 treatments performed with a robotic linear accelerator (CyberKnife), all SBRT was delivered using 6 to 18 MV photons delivered by standard clinical linear accelerators; for details on treatment planning, patient immobilization, and image guidance, see the references given above. Only patients with stage I disease according to the seventh lung cancer TNM classification and staging system[22] with complete information on prescribed PTV dose and number of fractions and not lost to follow-up within 1 month were considered for this analysis. Follow-up was defined as the time interval between the start of SBRT and local failure or censoring, respectively. This resulted in 875 patients treated for 940 lesions from the ECLRG database and 559 patients treated for 560 lesions from the DEGRO database. Regular computed tomography scans were performed during follow-up, and local recurrence was defined as radiologic disease progression in the treated lung parenchyma.

Three PTV dose-volume histogram parameters per SBRT were considered in this study: (1) dose prescribed to the PTV periphery (the near-minimum dose D_{min}); (2) the dose absorbed by 1% of the PTV (PTV D1% or near-maximum dose D_{max}); and (3) the average between D_{min} and D_{max} (D_{ave}) as a proxy for the mean dose received by the tumor. For 483 patients lacking PTV D1%, this dose parameter was estimated from the isocenter doses based on a linear regression equation derived from 147 treatments for which both PTV D1% and isocenter doses were available (Appendix EA.1, available online at https://doi.org/10.1016/j.ijrobp.2020.03.005).

Statistical analysis

For every patient i, D_{min}, D_{max}, and their average, D_{ave}, were converted to biologically effective doses (BEDs) according to

$$\mathrm{BED}_i = n_i D_i \left(1 + \frac{D_i}{\alpha/\beta}\right) \tag{1}$$

where n_i denotes the number of fractions, D_i the dose per fraction, and α/β is set to 10 Gy.

Collinearity among the 3 dose parameters (BED_{min}, BED_{max}, BED_{ave}) was assessed using variance inflation factors (VIFs) and hierarchical clustering with Spearman's rank correlation coefficients (ρ) as the similarity measure.[23]

In the first analysis, the 3 BED parameters were then correlated to the risk of local recurrence through Cox regression using the full sample of 1500 treatments. A separate Cox model was fitted to each dose parameter and each possible combination of dose parameters. The best-fitting model was determined based on the Akaike information criterion with second-order bias correction (AICc). Differences in AICc measure the strength of evidence for one model over another, and models differing by more than 8 from the model with the lowest AICc were considered to have substantially less evidential support from the data.[24]

In a second analysis, we treated the prediction of local tumor control as a classification task that presupposes that there are 2 classes of tumors: one that remains locally controlled and one that will regrow. Toward this aim, we decided to assume tumors as controlled if they remained controlled after at least 4 years of follow-up; only 7 out of 117 local recurrences (6%) were recorded after 4 years. Patients lost to follow-up (censored) before 4 years were therefore not considered for this analysis, yielding a sample of 386 patients with 401 treated tumors, of which 117 had recurred. The resulting data set was thus more balanced with respect to the binary endpoint of local control and local failure than the full data set.

The main classification method was logistic regression (LR), which estimates for each treated tumor a TCP by fitting

$$y_i = \frac{\exp\left(b_0 + \boldsymbol{x}_i^T \boldsymbol{b}\right)}{1 + \exp\left(b_0 + \boldsymbol{x}_i^T \boldsymbol{b}\right)} \tag{2}$$

where $y_i = 1$ if the tumor was controlled for patient i or $y_i = 0$ otherwise, and $\boldsymbol{x}_i^T \boldsymbol{b} = \sum_{j=1}^{p} x_{ij} b_j$ denotes the scalar product between the dose vector for patient i (consisting of p dose parameters, $p \in \{1, 2, 3\}$) and the corresponding vector of regression coefficients $\boldsymbol{b} = (b_1, \ldots, b_p)$. Note that for univariable analysis $x_{ij} = x_{i1} = BED_i$, and the regression coefficients can be related to the well-known TCP model parameters introduced by Okunieff et al[25] through

$$\mathrm{TCD50} = -\frac{b_0}{b_1} \tag{3}$$

and

$$k = \frac{1}{b_1} \tag{4}$$

As in Cox regression, the various LR classification models were also compared using the AICc[24] and by their (balanced) prediction accuracy (the average between sensitivity and specificity). Sensitivity denotes the probability of classifying a tumor as locally controlled given that it remains locally controlled, and specificity denotes the probability of classifying a tumor as recurrent given that it will regrow.

In a third analysis, we performed variable selection using 2 further classification algorithms: (1) LR with the least absolute shrinkage and selection operator (LASSO) method, which shrinks regression coefficients of unimportant variables to 0[26]; and (2) fast and frugal trees (FFTs). FFTs are decision trees with exactly 2 branches extending from each node, and either one or both branches are an exit branch leading to a classification decision.[27] This makes them easily interpretable and potentially attractive for clinical decision-making.

For all classification models, the test performance was estimated by randomly splitting the data set into 2 equal-sized parts, of which one was used for training the classifier and the other was used as a test set. This was repeated 100 times. Using LASSO and FFTs, the relative "importance" of the 3 dose parameters was estimated as the frequency with which each dose parameter was selected into a trained model and used for classification.

All analyses were performed in R version 3.5.0 with the glmnet package for building LASSO models and the FFTrees package for building FFTs. The latter were constructed by maximizing the weighted accuracy given by

$$w \times \text{sensitivity} + (1 - w) \times \text{specificity}$$

with $w = 0.4$ at each feature selection level because in practice specificity (correct classification of local failure) could be judged as more important than sensitivity. The "ifan" algorithm, which assumes independence among the features, was used for feature selection because it resulted in more intuitively interpretable trees than the "dfan" algorithm, which assumes dependencies among the features. The final tree was selected from a fan of 4 trees as the one having the highest weighted accuracy.

Statistical significance was defined as P values $<.005$.[28]

Results

General sample description

There were 1500 treatments in 1434 patients, of which 117 tumors did not remain locally controlled after a median follow-up of 20.7 months (range, 0.6-139 months). The median time to development of local recurrence was 15 months (range, 0.6-76.3 months). Actuarial local control rates at 12, 24, and 36 months were 96.8% (95% confidence interval, 95.8%-97.8%), 91.4% (89.7%-93.2%), and 89.0% (87.0%-91.1%), respectively. Characteristics of the full sample and the subset of treatments used for classification are given in Table 1.

Plotting D_{ave} against the planned gross tumor volume (GTV) mean dose (GTV Dmean, Appendix EA.2, available online at https://doi.org/10.1016/j.ijrobp.2020.03.005) for the subset of our data for which the latter was known, we could show that both were highly correlated according to GTV Dmean $= 0.207 + 1.096 \times D_{ave}$. Using correlation coefficients, the GTV Dmean was closely correlated with D_{ave} ($\rho = 0.931$) and D_{max} ($\rho = 0.927$), but not with D_{min} ($\rho = 0.714$). There was also a strong correlation between the BED based on the GTV Dmean and BED_{ave} ($\rho = 0.962$; Appendix EA.3, available online at https://doi.org/10.1016/j.ijrobp.2020.03.005), and their relation could be described by the formula $BED_{GTV\_Dmean} = -7.838 + 1.237 \times BED_{ave}$, showing that the BED based on the average between D_{min} and D_{max} tended to systematically overestimate the BED based on the GTV Dmean by $\approx 8\ Gy_{10}$.

Cox regression modeling results

The correlation between BED_{min} and BED_{max} in the complete sample of 1500 SBRT treatments was $\rho = 0.78$, and it was 0.92 and 0.95 for correlations between BED_{min} and BED_{ave} and BED_{max} and BED_{ave}, respectively. Cluster analysis classified BED_{max} and BED_{ave} into the same cluster, distinct from BED_{min}. These correlations were consistent with the VIFs, which indicated moderate collinearity between BED_{min} and BED_{max} but substantial collinearity between all other dose parameter combinations. In Cox regression, the best fit was obtained with a univariable model based on BED_{ave} (AICc = 1505.0), followed by the multivariable model including both BED_{ave} and BED_{min} (AICc = 1506.59) and the model including BED_{ave} and BED_{max} (AICc = 1506.63). There was no substantial evidential difference among the univariable and multivariable Cox models except for the univariable Cox

Table 1 Baseline characteristics of our sample used for Cox regression and classification*

Characteristic	Complete sample (N = 1500)	Subset with local control at least until 4 y (N = 284)	Subset with local failure (N = 117)	P value
Age, y	75 (31-93)	75 (31-93)	75 (50-88)	.697
Sex	Male: 855 (57%) Female: 645 (43%)	Male: 145 (51%) Female: 139 (49%)	Male: 67 (57%) Female: 50 (43%)	.273
Tumor size, cm	2.2 (0.5-5.0) (missing for 235)	1.9 (0.7-5.0) (missing for 17)	2.5 (1.0-4.8) (missing for 17)	4.802×10^{-5}
D_{min}, Gy	48 (12-64)	54 (12-60)	48 (19.2-64)	4.208×10^{-9}
BED_{min}, Gy_{10}	105.6 (16.8-180)	132 (16.8-180)	105.3 (43.2-180)	4.404×10^{-15}
D_{max}, Gy	66 (19.3-95.6)	68.6 (23.2-89.9)	62 (25.1-95.6)	1.464×10^{-10}
BED_{max}, Gy_{10}	178.5 (50-377)	205.3 (77-353)	156.6 (70-326.5)	1.122×10^{-13}
D_{ave}, Gy	56.8 (17.8-76.5)	60.6 (20.4-75)	54.2 (22.2-72.6)	9.934×10^{-11}
BED_{ave}, Gy_{10}	138.3 (49-269.7)	166.7 (56.1-241.9)	129.7 (55.9-248)	2.326×10^{-14}

Abbreviation: BED = biologically effective dose.

* Differences between continuous and categorical variables of tumors that remained locally controlled and recurrent, respectively, were tested using the Wilcoxon rank sum test and Fisher's exact test, respectively.

model based on BED_{min}, which had AICc = 1516.04, rendering it substantially worse than all other models.

Logistic regression and FFT classification results

In the data set of 401 treatments used for classification, the Spearman's rank correlation coefficients among the 3 BED parameters and their VIFs in the LR models were comparable to those reported for Cox regression on the complete data set. Accordingly, in the LR model containing BED_{min} and BED_{max}, both variables were significantly associated with TCP, but if BED_{min} or BED_{max} were included in a model in addition to BED_{ave}, their regression coefficient standard errors became inflated and P values became correspondingly large. The multivariable LR models yielded AICc values of 421.1 to 421.3 (2 dose parameters) and 423.0 (3 dose parameters). Characteristics of the univariable LR classification model fits are given in Table 2, and their predicted TCP curves as a function of BED are plotted in Figure 1. The AICc values show that the model based on the average dose (BED_{ave}) was preferred over all the other models, with substantially more evidence in its favor compared with the model based on BED_{max}. The classification performance measures in Table 2 assume that a tumor will be controlled if its TCP is ≥70%. The 70% threshold was chosen as a compromise in achieving both high sensitivity and specificity (ie, it was close to the threshold maximizing accuracy of the individual LR classifiers). Given this threshold, the model based on BED_{ave} also achieved the highest classification accuracy. For an alternative threshold of TCP ≥90%, the models predicted that BED_{ave} >196 Gy_{10} and BED_{max} >255 Gy_{10} would have to be prescribed to the PTV.

The FFT maximizing weighted accuracy with a sensitivity weight of 0.4 is depicted in Figure 2. This tree used all dose parameters for classification and implied that BED_{max} was the best predictor of local failure. The sensitivity, specificity, and accuracy achieved on the training data were 51.8%, 86.3%, and 69.0%, respectively.

Table 3 shows the estimated test performance of all models that was obtained after splitting the data set into a training and test set 100 times and averaging the results. The test performance was evaluated for 2 sensible TCP thresholds separately (70% and 90%) to show the effects on sensitivity and specificity. The results show that selecting a lower TCP threshold of 70% compared with 90% results in a much higher rate of correctly classified tumors that were locally controlled (sensitivity) and higher overall accuracy. At the 70% classification threshold, the model based on D_{min} had a significantly lower classification accuracy than that based on D_{max} ($P = 6.722 \times 10^{-8}$) or D_{ave} ($P < 2.2 \times 10^{-16}$). At the 90% threshold, however, the difference in model accuracy among the 3 univariable dose parameter models was not significant (LR model based on D_{min} vs D_{max}: $P = .375$; D_{min} vs D_{ave}: $P = .039$). The FFTs achieved comparable accuracy to the LR models, although they had a significantly lower area under the curve.

The majority of LASSO models (93%) included 2 dose parameters, with the importance of predictors being BED_{ave} (selected into 92% of models) before BED_{min} (85%) and BED_{max} (18%). For FFTs, the importance of predictors was BED_{max} (selected into 100% of the trees) before BED_{ave} (99%) and BED_{min} (66%).

Discussion

In SBRT, there can be substantial heterogeneity in the PTV dose prescription and distribution, with dose differences between the isocenter and PTV periphery of up to 50%. It is therefore not possible to know a priori which reported dosimetric parameter will best describe the dose–effect relationship. In this work, we investigated the importance of the 3 most relevant dosimetric parameters within the context of SBRT: (1) the dose prescribed to the isodose surface encompassing the PTV or near-minimum dose (D_{min}), which is recommended by ICRU report 91[10]; (2) the PTV D1% or near-maximum dose (D_{max}); and (3) the average between D_{min} and D_{max} (D_{ave}).

Previous meta-analyses based on LR modeling and including both conventionally and hypofractionated treatments of early-stage NSCLC have established dose–response relationships for both D_{min}[19,29] and D_{max}[19,30,31] after conversion into BEDs, as well as D_{ave} after conversion into equivalent doses in 2-Gy fractions.[32] In this work, we correlated these dose parameters after conversion into BEDs with the risk of local recurrence using Cox regression and to TCP using logistic regression and FFTs. We thereby found that most generally BED_{ave} was the best predictor of tumor response. In Cox regression, there was also evidence that the model based on BED_{min} resulted in a substantially worse fit to the data than

Table 2 Characteristics of the univariable logistic dose-response models*

Model parameter	TCD_{50}, Gy	k, Gy	AICc	Sensitivity†, %	Specificity†, %	Accuracy†, %	BED for 70% TCP, Gy_{10}
BED_{min}	50.9	29.7	423.7	79.4	61.0	70.2	77
BED_{max}	59.2	55.2	428.9	74.9	64.0	69.4	106
BED_{ave}	62.6	38.0	420.0	75.3	68.0	71.6	95

Abbreviations: AICc = Akaike information criterion with second-order bias correction; BED = biologically effective dose; TCP = tumor control probability.

* TCD_{50} and k are the TCP model parameters that can be computed from the logistic regression coefficients b_0 and b_1 (Equations 3 and 4).

† The classification performance is based on a probability threshold of 70% for classifying a tumor as locally controlled.

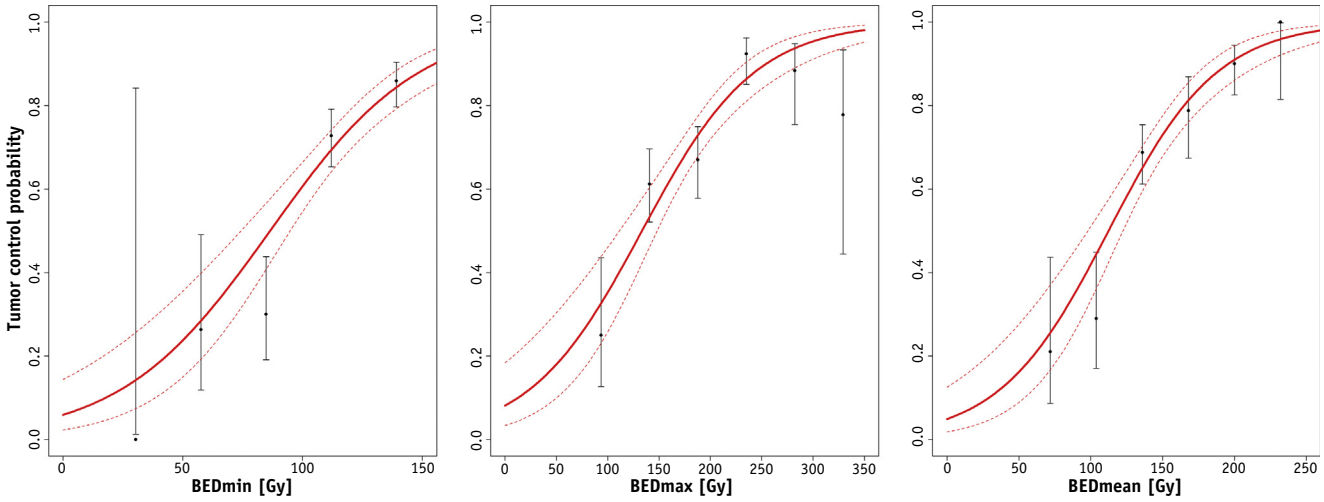

Fig. 1. Visualization of the logistic TCP models given in Table 2. There is 1 TCP point for each of 6 equally spaced bins along the biologically effective dose axis, and each TCP point was calculated by dividing the number of locally controlled tumors by the total number of tumors treated with a biologically effective dose falling within a given bin. Vertical bars represent Bayesian binomial 95% confidence intervals estimated according to Cameron.[43] *Abbreviation:* TCP = tumor control probability.

models based on the other dose parameters or combinations thereof. BED_{min} was also the least important predictor in the FFT model (Fig. 2). In the LR model, however, the worst fit was not obtained for BED_{min} but for BED_{max}. One explanation for this result could follow from the collinearity among the 3 dose parameters. Hierarchical cluster analysis and VIFs revealed that although BED_{min} and BED_{max} were only moderately correlated, there was high collinearity among all other BED combinations. Collinearity explains why adding more dose parameters to BED_{ave} did not improve model performance and why performance was not clearly superior for one parameter over the others. One possible way to deal with collinearity is to combine the correlated variables into a single parameter,[33] which is what we effectively did when creating D_{ave} as the average between D_{min} and D_{max}. This could explain why models based on BED_{ave} had a higher maximum likelihood and hence smaller AICc values than those based on either BED_{min} or BED_{max}. Another possibility is to build clusters of variables and

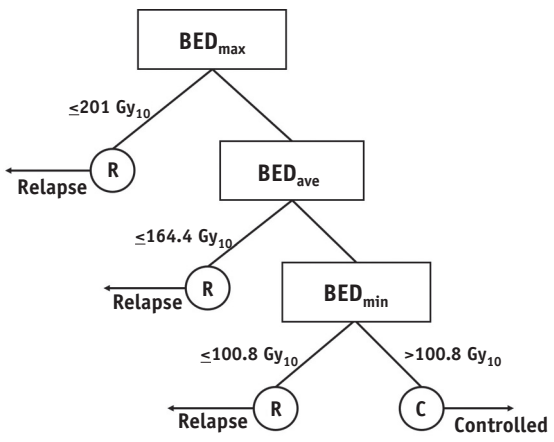

Fig. 2. Fast and frugal tree trained on 401 stereotactic body radiation therapy outcomes with 117 local relapses.

select 1 representative variable from each cluster.[34] It follows that aside from BED_{min}, either BED_{ave} or BED_{max} could be chosen as the second variable for making predictions of TCP. This is what the LASSO method has effectively done.

The estimated LR test performance (ie, the performance expected when applied to a new data set) depended on the choice of the TCP threshold used for classification. The optimal threshold at which the maximum accuracy was achieved varied among the different dose parameters. However, in practice one does not know the optimal threshold and has to decide on a certain probability cutoff. Our choices of 70% and 90% represent 2 possible choices for assuming a tumor is locally controlled. The 70% threshold was close to the TCP threshold at which the LR models achieved the highest accuracy when trained and applied to the complete data set. In both cases, BED_{ave} yielded the highest accuracy.

Our results somewhat contradict a previous analysis on a subset of the ECLRG database in which a better correlation between local relapse and BED_{min} compared with BED_{ave} or BED_{max} was observed in receiver operating characteristics analysis.[35] However, this previous analysis only included 26 patients with local relapse, so the robustness of its findings could be questioned. Modeling studies performed on the DEGRO database[17] and a collection of published study results[19] showed that SBRT isocenter doses (approximately BED_{max}) were better correlated with TCP than BED_{min}. Furthermore, a meta-analysis from 2012 including 15 SBRT studies found no significant correlation between the PTV prescription dose and TCP of early-stage NSCLC.[36] We have previously argued that isocenter doses might be more representative of the doses actually absorbed by the tumor, and the dose prescribed to the PTV periphery might be a suboptimal surrogate for the "true" therapeutic dose given the SBRT-typical inhomogeneity of dose

Table 3 Estimated test performance of the different classifiers

Model	AUC	Sensitivity at 90% TCP	Specificity at 90% TCP	Accuracy at 90% TCP	Sensitivity at 70% TCP	Specificity at 70% TCP	Accuracy at 70% TCP
LR: BED_{min}	0.737 ± 0.027	20.9 ± 21.6	92.0 ± 8.8	56.4 ± 6.5	55.4 ± 6.6	72.8 ± 5.9	64.1 ± 2.6
LR: BED_{max}	0.735 ± 0.028	18.5 ± 8.7	93.1 ± 3.7	55.8 ± 3.0	63.9 ± 5.1	68.6 ± 7.6	66.3 ± 2.9
LR: BED_{ave}	0.740 ± 0.027	24.2 ± 11.0	91.8 ± 4.6	58.0 ± 3.6	59.6 ± 5.9	75.0 ± 6.7	67.3 ± 2.2
LR: BED_{min} + BED_{max}	0.737 ± 0.028	26.4 ± 14.7	90.8 ± 5.9	58.6 ± 4.8	58.6 ± 6.6	73.5 ± 7.4	66.1 ± 3.0
LR: LASSO	0.738 ± 0.028	18.7 ± 14.3	93.4 ± 6.0	56.1 ± 4.5	58.4 ± 7.4	73.8 ± 7.6	66.1 ± 3.0
FFTs*	0.677 ± 0.027	55.3 ± 7.5	80.1 ± 10.0	67.7 ± 2.9	55.3 ± 7.5	80.1 ± 10.0	67.7 ± 2.9

Abbreviations: AUC = area under the curve; BED = biologically effective dose; FFT = fast and frugal tree; LASSO = least absolute shrinkage and selection operator; LR = logistic regression; TCP = tumor control probability.

* Note that the FFTs produce a discrete output (probability either 0 or 1).

distribution within the GTV, the variability of safety margins and GTV sizes, and tumor movement, as well as uncertainties concerning the actual delivered dose owing to limitations of the adopted dose calculation algorithms used for treatment planning.[17] Along these lines, Bibault et al showed that PTV D95% doses calculated by a type A algorithm correlated only weakly with GTV Dmean calculated by a Monte Carlo algorithm, in particular for small GTVs (<20 cm^3).[37] We confirm these results because in our cohort the GTV Dmean was closely correlated with D_{ave} and D_{max}, but not with D_{min}.

An important clinical implication of our results is that, because doses at the PTV periphery are less accurate for predicting TCP, higher emphasis could be placed on achieving sufficiently high mean and maximum doses within the GTV rather than the PTV-covering dose. In other words, the GTV Dmean may be a more important prescription target than the PTV-encompassing dose. However, note that in the FFT depicted in Figure 2, all 3 dose parameters were required to exceed a particular threshold for a tumor to be classified as staying locally controlled. Furthermore, a recent study found that too little dose in a 30-mm shell around the PTV in SBRT for early-stage NSCLC was associated with an increased risk of distant metastasis,[38] which—if confirmed as a causal association—would emphasize the importance of sufficient dose to the PTV periphery and beyond for reasons other than local control.

As a major limitation of this study, a large number of data important from a radiobiological perspective were missing, in particular equivalent uniform doses[39] that might correlate even better to TCP than the dose parameters tested here, with the PTV D1% missing for 483 treatments that we had to estimate from isocenter doses and maximum tumor diameters missing for 235 tumors. Tumor size could be a confounder because it might be correlated to both dose (the cause) and tumor control (the effect). Maximum tumor diameters were significantly smaller in tumors that remained locally controlled (Table 1) and were positively correlated with D_{min} ($\rho = 0.080$, $P = .0045$), D_{max} ($\rho = 0.124$, $P = 9.17 \times 10^{-6}$), and D_{ave} ($\rho = 0.114$, $P = 4.77 \times 10^{-5}$). However, there was only a small and nonsignificant correlation between BED parameters and tumor diameter (all Spearman's $|\rho| \leq .03$, $P > .25$), because larger tumors tended to receive a larger number of fractions ($\rho = 0.048$, $P = .091$). When we restricted the analysis to tumors with known sizes and adjusted for maximum tumor diameter, the results did not change qualitatively.

Another limitation is that this study is only based on planned doses, not taking their uncertainties into account. Furthermore, converting doses to BED comes with a scaling effect that makes small dose differences in delivered dose appear larger. However, when we restricted the analysis to 727 SBRT treatments performed with 3 fractions only and used physical doses instead of BEDs, the average dose remained a better predictor of TCP than the PTV prescription dose (D_{min}) in both Cox and logistic regression, as shown in Appendix EA.4 (available online at https://doi.org/10.1016/j.ijrobp.2020.03.005).

Finally, our BED calculations are based on the linear-quadratic model with an α/β ratio of 10 Gy. Although the validity of this model has been questioned for SBRT,[40] analyses based on clinical data have not provided evidence that it should be abandoned.[41] Concerning the α/β ratio for SBRT of early-stage NSCLC, we have recently shown that $\alpha/\beta = 10$ Gy was consistent with individual local tumor control data, although somewhat higher estimates in the range of 12 to 16 Gy yielded better model fits[42]; nevertheless, we here decided to use $\alpha/\beta = 10$ Gy to ease the comparison of our BED thresholds with past studies. We emphasize that because of these limitations, our results need validation in different data sets.

Conclusions

BEDs based on the average between near-maximum and near-minimum doses (BED_{ave}) were better correlated with TCP than either of these dose parameters alone, although the evidence supporting BED_{ave} as a superior predictor of TCP was not substantial except compared with BED_{min} in the Cox model and BED_{max} in logistic regression. Because D_{ave} was highly correlated to GTV Dmean as both physical dose and BED, the latter may be used as a prescription target aside from the PTV-encompassing dose and should yield a BED of at least 150 Gy_{10} to achieve a TCP exceeding 90%.

References

1. Guckenberger M, Andratschke N, Alheit H, et al. Definition of stereotactic body radiotherapy: Principles and practice for the treatment of stage I non-small cell lung cancer. *Strahlenther Onkol* 2014;190:26-33.
2. Alayed Y, Cheung P, Vesprini D, et al. SABR in high-risk prostate cancer: Outcomes from 2 prospective clinical trials with and without elective nodal irradiation. *Int J Radiat Oncol Biol Phys* 2019;104:36-41.
3. Brunner TB, Blanck O, Lewitzki V, et al. Stereotactic body radiotherapy dose and its impact on local control and overall survival of patients for locally advanced intrahepatic and extrahepatic cholangiocarcinoma. *Radiother Oncol* 2019;132:42-47.
4. Tree AC, Khoo VS, Eeles R, et al. Stereotactic body radiotherapy for oligometastases. *Lancet Oncol* 2013;14:e28-e37.
5. Lewis SL, Porceddu S, Nakamura N, et al. Definitive stereotactic body radiotherapy (SBRT) for extracranial oligometastases. *Am J Clin Oncol* 2017;40:418-422.
6. Klement RJ, Abbasi-Senger N, Adebahr S, et al. The impact of local control on overall survival after stereotactic body radiotherapy for liver and lung metastases from colorectal cancer: A combined analysis of 388 patients with 500 metastases. *BMC Cancer* 2019;19:173.
7. Videtic GMM, Donington J, Giuliani M, et al. Stereotactic body radiation therapy for early-stage non-small cell lung cancer: Eexecutive summary of an ASTRO evidence-based guideline. *Pract Radiat Oncol* 2017;7:295-301.
8. Guckenberger M, Andratschke N, Dieckmann K, et al. ESTRO ACROP consensus guideline on implementation and practice of stereotactic body radiotherapy for peripherally located early stage non-small cell lung cancer. *Radiother Oncol* 2017;124:11-17.
9. Schneider BJ, Daly ME, Kennedy EB, et al. Stereotactic body radiotherapy for early-stage non-small-cell lung cancer: American Society of Clinical Oncology endorsement of the American Society for Radiation Oncology evidence-based guideline. *J Clin Oncol* 2018;36:710-719.
10. Seuntjens J, Lartigau EF, Cora S, et al. ICRU report no. 91: Prescribing, recording, and reporting of stereotactic treatments with small photon beams. *J ICRU* 2014;14.
11. Moustakis C, Blanck O, Ebrahimi Tazehmahalleh F, et al. Planning benchmark study for SBRT of early stage NSCLC: Results of the DEGRO working group Stereotactic Radiotherapy. *Strahlenther Onkol* 2017;193:780-790.
12. Daly ME, Perks JR, Chen AM. Patterns-of-care for thoracic stereotactic body radiotherapy among practicing radiation oncologists in the United States. *J Thorac Oncol* 2013;8:202-207.
13. Guckenberger M, Allgäuer M, Appold S, et al. Safety and efficacy of stereotactic body radiotherapy for stage I Non-small-cell lung cancer in routine clinical practice: a patterns-of-care and outcome analysis. *J Thorac Oncol* 2013;8:1050-1058.
14. Corso CD, Park HS, Moreno AC, et al. Stage I lung SBRT clinical practice patterns. *Am J Clin Oncol* 2017;40:358-361.
15. Wilke L, Andratschke N, Blanck O, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams. *Strahlenther Onkol* 2019;195:193-198.
16. Ohri N, Werner-Wasik M, Grills IS, et al. Modeling local control after hypofractionated stereotactic body radiation therapy for stage I non-small cell lung cancer: a report from the Elekta collaborative lung research group. *Int J Radiat Oncol Biol Phys* 2012;84:e379-e384.
17. Guckenberger M, Klement RJ, Allgäuer M, et al. Applicability of the linear-quadratic formalism for modeling local tumor control probability in high dose per fraction stereotactic body radiotherapy for early stage non-small cell lung cancer. *Radiother Oncol* 2013;109:13-20.
18. Tai A, Liu F, Gore E, et al. An analysis of tumor control probability of stereotactic body radiation therapy for lung cancer with a regrowth model. *Phys Med Biol* 2016;61:3903-3913.
19. Santiago A, Barczyk S, Jelen U, et al. Challenges in radiobiological modeling: Can we decide between LQ and LQ-L models based on reviewed clinical NSCLC treatment outcome data? *Radiat Oncol* 2016;11:67.
20. Liu F, Tai A, Lee P, et al. Tumor control probability modeling for stereotactic body radiation therapy of early-stage lung cancer using multiple bio-physical models. *Radiother Oncol* 2016;122:286-294.
21. Grills I, Hope A, Guckenberger M, et al. A Collaborative analysis of stereotactic lung radiotherapy outcomes for early-stage non—small-cell lung cancer using daily online cone-beam computed tomography image-guided radiotherapy. *J Thorac Oncol* 2012;7:542-551.
22. Mirsadraee S, Oswal D, Alizadeh Y, et al. The 7th lung cancer TNM classification and staging system: Review of the changes and implications. *World J Radiol* 2012;4:128-134.
23. Harrell FE Jr. *Regression Modeling Strategies*. 2nd ed. New York: Springer; 2015.
24. Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Res* 2004;33:261-304.
25. Okunieff P, Morgan D, Niemierko A, et al. Radiation dose-response of human tumors. *Int J Radiat Oncol Biol Phys* 1995;32:1227-1237.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.
27. Phillips ND, Neth H, Woike JK, et al. FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgm Decis Mak* 2017;12:344-368.
28. Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav* 2018;2:6-10.
29. Stuschke M, Pöttgen C. Altered fractionation schemes in radiotherapy. *Front Radiat Ther Oncol* 2010;42:150-156.
30. Mehta N, King CR, Agazaryan N, et al. Stereotactic body radiation therapy and 3-dimensional conformal radiotherapy for stage I non-small cell lung cancer: A pooled analysis of biological equivalent dose and local control. *Pract Radiat Oncol* 2012;2:288-295.
31. Shuryak I, Carlson DJ, Brown JM, et al. High-dose and fractionation effects in stereotactic radiation therapy: Analysis of tumor control data from 2965 patients. *Radiother Oncol* 2015;115:327-334.
32. Jeong J, Oh JH, Sonke J-J, et al. Modeling the cellular response of lung cancer to radiation therapy for a broad range of fractionation schedules. *Clin Cancer Res* 2017;23:5469-5479.
33. Midi H, Sarkar SK, Rana S. Collinearity diagnostics of binary logistic regression model. *J Interdiscip Math* 2010;13:253-267.
34. Sanche R, Lonergan K. Variable reduction for predictive modeling with clustering. *Casualty Actuarial Society Forum, Winter* 2006;89-100.
35. Kestin L, Inga G, Guckenberger M, et al. Dose-response relationship with clinical outcome for lung stereotactic body radiotherapy (SBRT) delivered via online image guidance. *Radiother Oncol* 2014;110:499-504.
36. van Baardwijk A, Tomé WA, van Elmpt W, et al. Is high-dose stereotactic body radiotherapy (SBRT) for stage I non-small cell lung cancer (NSCLC) overkill? A systematic review. *Radiother Oncol* 2012;105:145-149.
37. Bibault JE, Mirabel X, Lacornerie T, et al. Adapted prescription dose for Monte Carlo algorithm in lung SBRT: Clinical outcome on 205 patients. *PLoS One* 2015;10:1-10.
38. Diamant A, Chatterjee A, Faria S, et al. Can dose outside the PTV influence the risk of distant metastases in stage I lung cancer patients treated with stereotactic body radiotherapy (SBRT)? *Radiother Oncol* 2018;128:513-519.
39. Niemierko A. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Med Phys* 1997;24:103-110.
40. Shibamoto Y, Miyakawa A, Otsuka S, et al. Radiobiology of hypofractionated stereotactic radiotherapy: What are the optimal fractionation schedules? *J Radiat Res* 2016;57:i76-i82.
41. Brown JM. The biology of SBRT: LQ or something new? *Int J Radiat Oncol Biol Phys* 2018;101:964.
42. Klement RJ, Sonke JJ, Allgäuer M, et al. Estimation of the α/β ratio of non-small cell lung cancer treated with stereotactic body radiotherapy. *Radiother Oncol* 2020;142:210-216.
43. Cameron E. On the estimation of confidence intervals for binomial population proportions in astronomy: The simplicity and superiority of the Bayesian approach. *Publ Astron Soc Aust* 2011;28:128-139.

Appendix A

A.1 Estimation of PTV D1% from isocenter dose

All SBRT treatments from the xxx database had isocenter doses available, but PTV D1% was only available for a subset of 79 SBRT treatments from two institutions (Schweinfurt and Güstow). In contrast, the DVH parameters PTV D1% and GTV mean dose were available for all treatments within the ECLRG database. We therefore used the 79 treatments from the DEGRO database and 68 treatments from the ECLRG database from the University hospital of Würzburg for which both PTV D1% and isocenter doses were available in order to derive a linear regression equation which was used to estimate PTV D1% doses for all DEGRO treatments lacking DVH parameters (Figure A1).

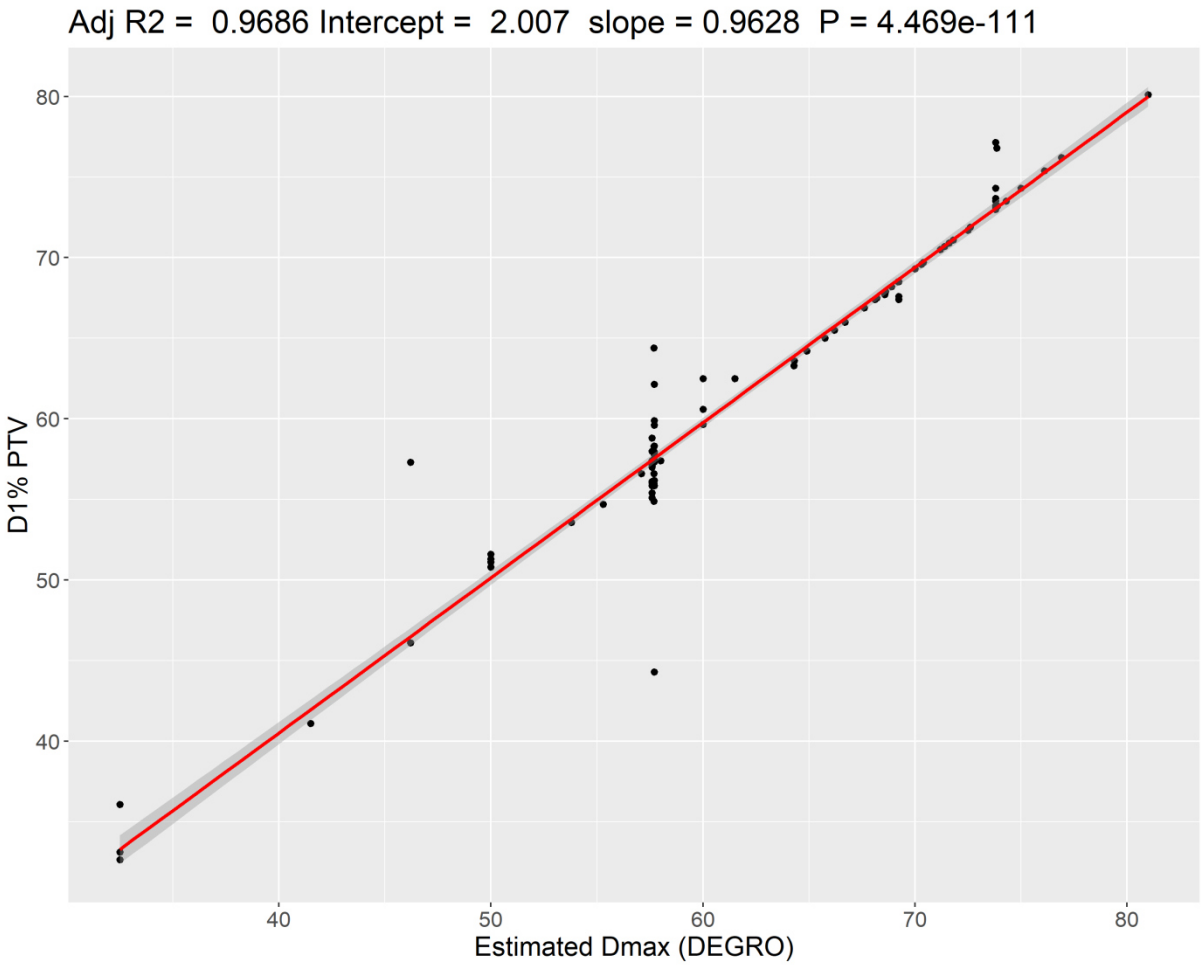

Figure A1: Calibration line for transforming D_{max} in the DEGRO database to a D1% PTV parameter. The transformation is based on 68 treatments from the University Hospital of Würzburg (Germany) which are shared by both databases, 20 treatments from the Leopoldina Hospital in Schweinfurt (Germany)

and 59 treatments from the Cyberknife center in Güstrow (Germany) for which both variables were available.

A.2 Estimation of GTV mean dose from isocenter and prescribed dose

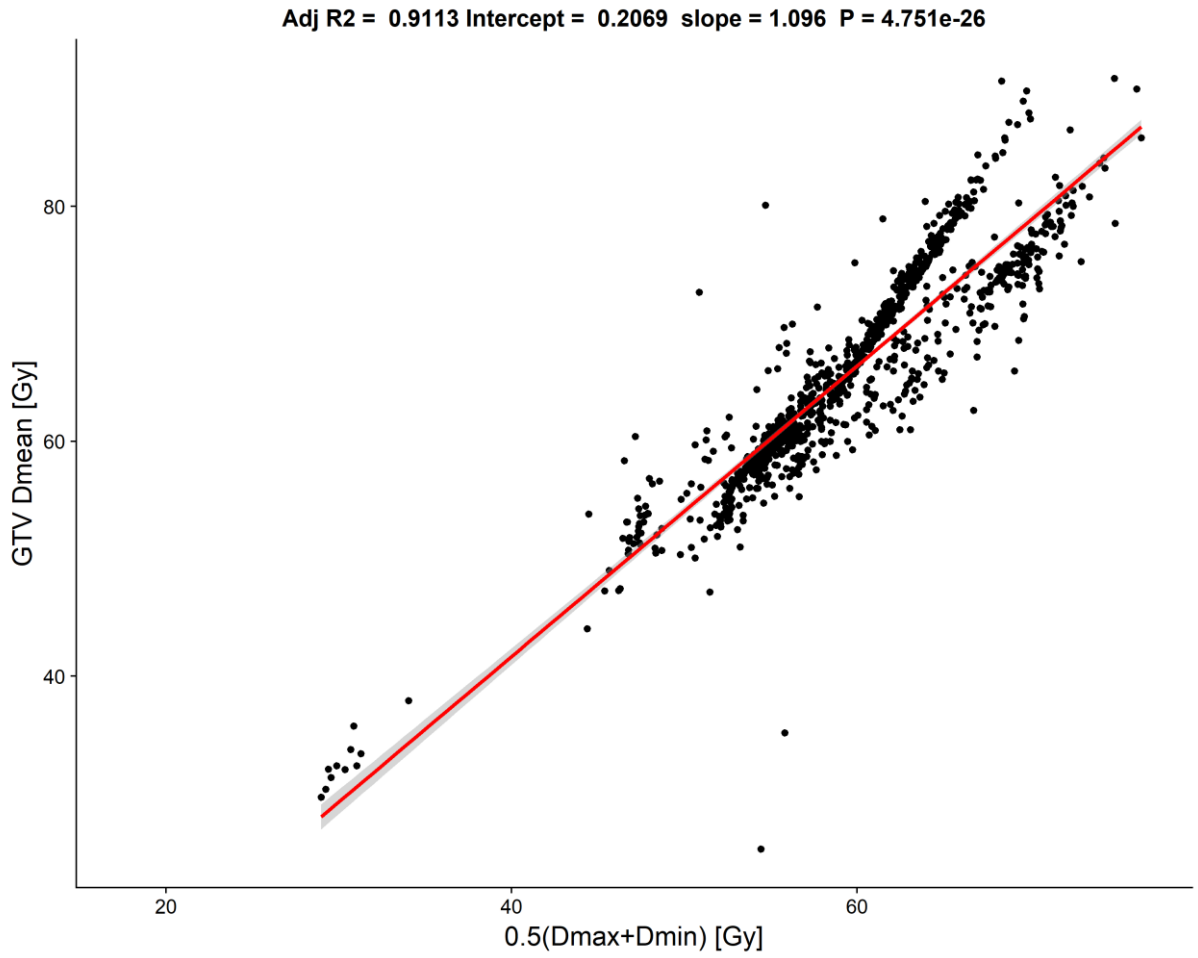

Figure A2: Correlation between D_{mean} and the GTV mean dose based on 940 treatments from the ECLRG database, 59 patients from the Cyberknife center in Güstrow (Germany) and 18 treatments from the Leopoldina Hospital Schweinfurt (Germany) for which GTV mean dose, PTV 1% and the prescribed dose to the PTV periphery were known.

A.3 Estimation of BED based on GTV mean dose from BED_{ave}

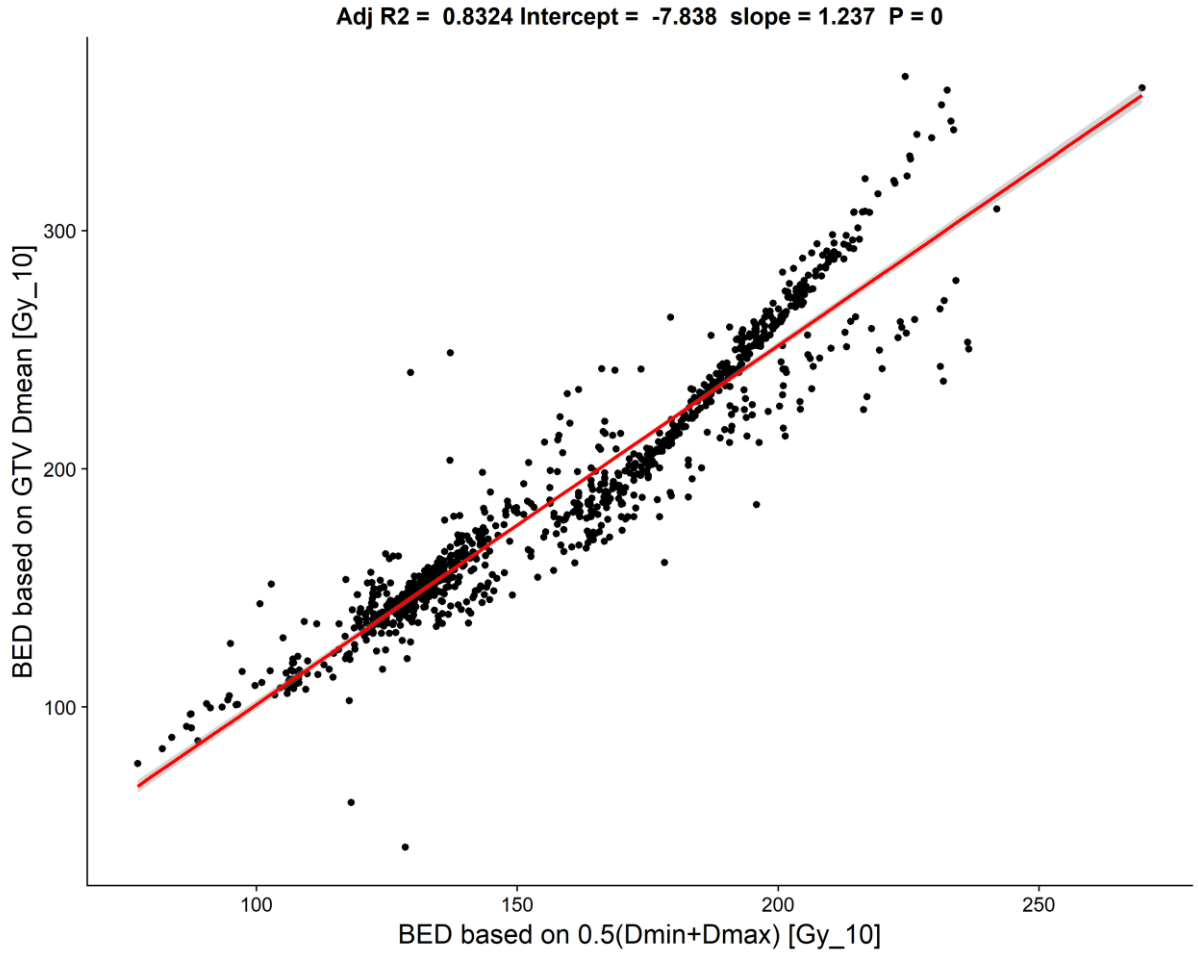

Figure A3: Correlation between BED_{ave} and the BED based on the GTV mean dose adopting an α/β ratio of 10 Gy. The figure is based on the same subset of the data as Figure A2.

A.4 Analysis restricted to physical doses delivered in three fractions

We performed a re-analysis on the subgroup of patients having received three fractions, so that physical doses instead of BEDs could be used for modelling. In total, 727 SBRTs were performed with three fractions; median prescription doses (D_{min}) and PTV D1% (D_{max}) were 54 Gy (range 12−60 Gy) and 68.6 Gy (32.9−92.4 Gy), respectively., resulting in a median D_{ave} of 59.0 Gy (28.6−76.2 Gy).

Building a multivariable Cox model from all three dose parameters with the LASSO method resulted in D_{max} and D_{ave} being selected into the model, but not D_{min}. While the univariable Cox model based on D_{max} showed the best fit according to the AICc (378.3), it performed almost equal to the model based on D_{ave} (AICc=378.6), and both models had somewhat more evidential support than the model based on D_{min} (AICc=381.7). There was no substantially better support for multivariable models compared to the univariable models. Building logistic regression models for classifying 33 local recurrences and 159 tumors that remained locally controlled over at least 4 years, D_{ave} (AICc=164.7) was a better predictor than either D_{max} (AICc=167.0) or D_{min} (AICc=167.4). The LASSO method selected only D_{ave} into a potentially multivariable logistic regression model. Finally, a fast and frugal tree built on all three dose parameters used only D_{max} and D_{ave} for deciding between local control and recurrence (Figure A4).

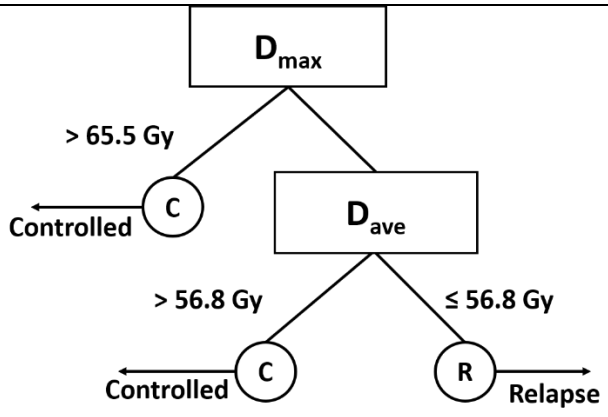

Figure A4: Fast and frugal tree fitted to a subset of tumors which had been treated with three fractions and had either locally recurred (N=33) or stayed controlled for at least four years of follow-up (N=159).