Dose-escalation in lung cancer

Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients

Steven F. Petit\textsuperscript{a,*}, Wouter van Elmpt\textsuperscript{b}

\textsuperscript{a}Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam; and \textsuperscript{b}Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, The Netherlands

\textbf{A R T I C L E  I N F O}

Article history:
Received 6 March 2015
Received in revised form 5 June 2015
Accepted 25 July 2015
Available online 22 August 2015

Keywords:
Treatment planning QA
Knowledge based treatment planning
Dose–distance relation
NSCLC dose escalation
Dose prediction

\textbf{A B S T R A C T}

\textbf{Purpose/objective:} To develop a method to predict feasible organ-at-risk (OAR) and tumour dose levels of non-small cell lung cancer (NSCLC) patients prior to the start of treatment planning.

\textbf{Materials/methods:} Included were NSCLC patients treated with volumetric modulated arc therapy according to an institutional isotoxic dose-escalation protocol. A training cohort (N = 50) was used to calculate the average dose inside the OARs as a function of the distance to the planning target volume (PTV). These dose–distance relations were used in a validation cohort (N = 39) to predict dose–volume histograms (DVHs) of OARs and PTV as well as the maximum individualized PTV dose escalation.

\textbf{Results:} The validation cohort showed that predicted and achieved MLD were in agreement with each other (difference: -0.1 ± 1.9 Gy, \(p = 0.81\)). The spinal cord was dose limiting in only two patients, which was accurately predicted. The achieved mean PTV dose varied from 52 to 73 Gy and was predicted correctly with an accuracy better than 2 Gy (i.e. 1 fraction) for 79% of the patients.

\textbf{Conclusion:} We have shown that the MLD and the prescribed PTV dose could be accurately predicted for NSCLC patients. This method can guide the treatment planner to achieve optimal OAR sparing and tumour dose escalation.

© 2015 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 117 (2015) 453–458
The goal of the current study was to develop an easy to implement model that predicts DVHs based solely on the anatomy and treatment plans of previously treated patients, and furthermore can rank a planned DVH compared to prior treatment plans. This method can then be used to determine the maximum dose escalation for advanced stage NSCLC lung patients. The method was investigated for a large cohort of non-small cell lung cancer (NSCLC) patients treated with a dose-escalation protocol.

Materials and methods

Patients and treatment

One hundred twenty-five stage II/III NSCLC patients from Maastro Clinic, The Netherlands, were treated in 2013 with concurrent chemo-radiotherapy according to an institutional isotoxic dose-escalation protocol [NCT01166204]. Data of ninety-five patients could be retrieved. Six patients were excluded because the dose constraints deviated from the standard protocol due to previous thoracic irradiation or surgery, simultaneous irradiation of a bone metastasis, or the presence of a pacemaker, leaving a total of 89 patients. Patients were treated using volumetric modulated arc therapy (Varian Rapid Arc) with 1 to 4 arcs (median of 2 arcs). The gross tumour volume (GTV) of the primary tumour and the involved lymph nodes was expanded with 5 mm to define the clinical target volume (CTV) that was allowed to be edited according to anatomy (e.g. bony anatomy). The planning target volume (PTV) was the union of the primary tumour CTV plus a margin of 10 mm in all directions and the CTV of the lymph nodes plus a margin of 5 mm. The prescribed dose to the PTV was 45 Gy delivered in 1.5 Gy per fraction bidaily and further escalated in fractions of 2 Gy once daily up to a maximum of 69 Gy while keeping within the OAR dose constraints. The dose to the PTV of the lymph nodes was escalated to the same level as the PTV of the primary tumour. The following planning objectives were used: 99% of the PTV must receive a dose of at least 90% of the prescribed dose, the mean dose in the combined lungs (MLD) excluding the GTV <20 Gy; maximum dose to plexus brachialis <70 Gy; mean dose to the heart <46 Gy; maximum dose to the heart, mediastinal envelop and esophagus <76 Gy; maximum dose to the at the most 0.1 cc (D0.1cc) of the spinal cord <54 Gy. The spinal cord was delineated from the most cranial to caudal slice of the planning CT.

For these patients, except for the spinal cord and lungs, none of the OARs were limiting for the maximum PTV prescription dose of 69 Gy. Therefore the analysis was focused on the lungs (excluding GTV) and spinal cord.

The DVH prediction model

We developed a prediction model with the underlying assumption that on average the dose to a voxel of an OAR decreases with increasing distance from the PTV. This was quantified using the dose–distance relation, defined as the average dose in the voxels of an organ as a function of the distance to the PTV. For the same treatment modality, beam configuration and beam energy the achievable dose–distance relation for a given OAR will be robust and almost patient independent. This dose–distance relation is referred to as the population dose–distance relation.

Once the population dose–distance relation for an OAR is available, for a specific patient and organ the expected dose to each voxel can be predicted when the distances from all voxels of the organ to the PTV are known. Next the expected DVH is calculated, and used as prediction of the achievable DVH. The distance of every voxel of an OAR to the PTV is calculated solely based on the contours. So once the contours for the OAR and PTV of a given patient are available the expected DVHs can be calculated.

Calculation of the dose–distance relation per patient

The dose–distance relation of the lungs-GTV and the spinal cord were calculated for 50 patients selected at random (training cohort). Because patients were treated with different prescribed doses, first the individual dose distribution was normalized to the mean PTV dose of that patient. Delineations of the OARs and PTV were extracted in DICOM RT format from the treatment planning system (Eclipse 11.0, Varian Medical Systems, Palo Alto, CA). Per CT slice the contours were rasterized in a binary mask with a resolution of $1 \times 1 \text{mm}^2$. The 2D masks were stacked in a 3D matrix. Next a marching cube algorithm was used to generate triangular mesh elements at the surface of the mask and assembled in a 3D mesh volume. From the mesh volumes, points were sampled automatically using a Hammersley sequence, that ensured a homogeneous spread of sample points in the volume with an average distance between points of 1.5 mm (yielding 296 points per cc). The dose at each point was derived using a tri-linear interpolation of the dose grid, which had a resolution of $2.5 \times 2.5 \times 2.5 \text{mm}^3$. For each point the distance to the mesh of the PTV was derived by calculating the distance from the point to each mesh triangular element and taking the minimum over the elements. For points of the OAR within the PTV, the distance to the surface of the PTV was set negative. The data were binned based on the distance of the PTV, with bins of 1 mm width. For all points in each bin the mean dose was calculated. The dose–distance relation for a given patient and OAR was defined as the mean dose per bin as function of its distance from the PTV. See also Fig. 1. For the mathematical formulation and for an alternative strategy to calculate the dose–distance relation, that can be performed using standard volume operations such as subtractions, unions and intersections see Supplementary material A.

The population dose–distance relations

For a specific organ and treatment technique, all patients will have similar dose–distance relations. However, due to differences in plan quality among patients, the dose–distance relations are likely to vary. The steeper the dose fall off of the dose–distance relation the better the organ was spared. To predict DVHs of new patients, in this study three population dose–distance relations were used per organ: the 25th, 50th (median) and 75th percentiles of the dose–distance relations over all patients in the training cohort. The percentiles were calculated for each distance bin separately. DVHs predicted based on these population dose–distance relation correspond to the plans of the best 25%, 50% and 75% of the patients in the training cohort, respectively.

Prediction of DVH parameters

The remaining 39 patients (validation cohort) were used to validate the model. From the OARs of each patient in the validation cohort, calculation points were sampled as described above. For each point the distance to the surface of the PTV was calculated. Next, the normalized dose to each calculation point was estimated by looking up the dose corresponding to that distance from the population dose–distance relations. The expected DVH, DVHnorm, was calculated by calculating the inverse cumulative histogram of the doses in the calculation points. This was repeated for the three population dose–distance relations. For the spinal cord the D0.1cc, norm and for the lungs the mean lung dose (MLDnorm) were derived from the corresponding DVHnorm.
Both the spinal cord and lungs could be dose limiting in the dose escalation protocol. The achievable PTV dose will depend on the first OAR for which the constraints value is reached. The maximum allowed $D_{0.1cc}$ of the spinal cord was 54 Gy and the maximum allowed MLD was 20 Gy. So by dividing the constraint values by the predicted normalized values (thus $PTV_{\text{mean, lungs}} = 20 \text{ Gy}/\text{MLD}_{\text{norm}}$ and $PTV_{\text{mean, spinal cord}} = 54 \text{ Gy}/D_{0.1cc}$, norm), the mean PTV dose was predicted for each organ separately. The prediction of the mean PTV dose, $PTV_{\text{mean}}$, was the minimum of $PTV_{\text{mean, lungs}}$ and $PTV_{\text{mean, spinal cord}}$.

Comparison between the predictions and the achieved DVHs

The expected $D_{0.1cc}$, norm, the MLD, and the PTV$\text{mean}$ were compared with the achieved values for the patients from the validation cohort. To interpret the prediction errors of the model also in units of gray, in addition the expected, non-normalized, $D_{0.1cc}$ and MLD were compared with the achieved $D_{0.1cc}$ and the MLD. The latter were calculated by multiplying the expected $D_{0.1cc}$, norm and MLD$\text{norm}$ with the achieved $PTV_{\text{mean}}$. The mean, standard deviation (SD), minimum and maximum of the differences were calculated. The Pearson correlation coefficient (Pearson’s $r$) was calculated for the expected and achieved values and a paired $t$-test was used to check for statistically differences between the expected and achieved values. A Pearson’s $r$ of 1 indicates a perfect correlation. A $p$-value of 0.05 was used to test for statistical significance. Differences are indicated as mean ± 1 SD (range: minimum–maximum; $p$-value; Pearson’s $r$). For comparison the MLD, $D_{0.1cc}$ and mean target dose were also predicted and plotted for the patients from the training cohort. All calculations were performed using Erasmus MC RTStudio 1.24 and Python 2.7.

Results

The training and validation group consisted of 50 and 39 patients, respectively. The mean PTV dose varied from 53.8 Gy to 72.5 Gy in the training cohort and from 51.9 Gy to 72.9 Gy in the validation cohort. For 76% and 67% of the patients the maximum prescribed dose escalation threshold of 69 Gy was reached in the training and validation cohort, respectively.

Fig. 1(b) shows the relation between the distance from the PTV and the dose of the sampled lungs points for one patient. The variation of the dose as function of the distance is considerable but on average the dose decreases with increasing distance the lungs, also for the spinal cord (not shown). The dose–distance relations of all patients in the training group are shown in Fig. 2. Also the 25th, 50th and 75th population dose–distance relations of all training patients combined are indicated. For the lungs the average dose fall off within the first 3 cm of the PTV is similar for all patients. As the distance increases the variation increases as well. For the spinal cord the dose–distance relations are less smooth than for the lungs.

The predicted MLD of the validation patients are shown in Fig. 3. For comparison the predictions of the training data were added as well. For the validation data the average difference between the achieved and expected MLD normalized to the mean PTV dose was $-0.0019 ± 0.029$; $-0.066–0.068$; $p = 0.69$; Pearson’s $r = 0.94$). This would have translated into an average absolute difference in MLD of $−0.11 ± 1.9 \text{ Gy}$ ($−3.8–4.8$; $p = 0.71$; Pearson’s $r = 0.91$).

For the spinal cord the expected $D_{0.1cc}$ were on average considerably lower than the achieved values (normalized dose: mean: $-0.14 ± 0.15$ ($-0.52–0.12$; $p < 0.001$; Pearson’s $r = 0.63$) and absolute dose: mean: $−9.4 ± 9.8 \text{ Gy}$ ($-32.0–8.9 \text{ Gy}$; $p < 0.001$; Pearson’s $r = 0.63$). See also Fig. 4.

The predicted mean PTV doses are shown in Fig. 5. For 79% of the patients in the validation cohort the predicted dose was within 2 Gy (which is 1 treatment fraction) from the achieved dose and for 85% of the patients the achieved dose was within the uncertainty of the prediction, also indicated by the error bars in the figure. Since the maximum dose escalation in the clinical protocol was 69 Gy (prescribed dose), the prediction was considered correct if both the predicted and achieved mean PTV dose were at least 69 Gy. The MLD and $D_{0.1cc}$ were dose limiting for 8 and 2 patients from the validation cohort respectively, which was correctly predicted by the model for each case.

Discussion

In this study a method was developed to predict the OAR dose levels and the feasible amount of tumour dose-escalation (up to 69 Gy) for advanced stage NSCLC patients treated with volumetric arc therapy with an iso-toxic dose escalation protocol. DVHs are predicted based solely on the contours of the OAR and PTV and using a database of previously treated patients. The method accurately predicted the mean lung dose within $−0.11 ± 1.9 \text{ Gy}$ (mean ± SD) and the maximum achievable dose escalation to the PTV, within one 2 Gy fraction for 79% of the patients. The expected
and achieved MLD values were not statistically different and had a high correlation (Pearson’s $r = 0.91$).

Previous methods to evaluate plan quality can be divided into two categories. The first category contains models that make predictions based only on the distance and overlap of OARs with the PTV [1–5]. These models have shown to be robust for head-and-neck and prostate cancer treatments and they are relatively easy to implement. Methods from the second category use more sophisticated algorithms, such as deformable image registration or principal component analysis [6,8,13]. These models can predict dose and provide estimates of the prediction accuracy, but require more resources to implement clinically. Some have been successfully commercialized. The current method falls within the first category. This is the first time that such a method has been demonstrated for lung cancer patients, that can predict the maximum tumour dose escalation and that provides a confidence interval on the expected DVH parameters.

Confidence intervals are important, because due to heterogeneity in the treatment plans and the very basic model assumptions a predicted DVH value may be unachievable for a given patient. The confidence interval provided by our method, allows ranking an achieved DVH parameter to the percentage of previously treated patients for which higher and lower values were achieved. For instance if for a given patient the achieved MLD is 1 Gy higher than the expected MLD, but still corresponds to the better e.g. 25% of the previously treated patients, the difference between the expected value and achievable value may still be acceptable. This study focused on the MLD and $D_{0.1cc}$ as recommended by ICRU-83.

For the MLD the predicted values of the validation patients resembled the achieved values better than for the spinal cord $D_{0.1cc}$. This was expected, because for most patients in the training and validation set the constraint on the MLD was dose limiting. Since the spinal cord can be considered as a serial organ, the expected benefit of lowering the dose further than its constraint value is limited and was in clinical practice not aimed for. Hence the achieved $D_{0.1cc}$ values were considerably higher than the predicted values. This also explains why the variation in the dose–distance relation is larger for the spinal cord. However, for
both patients in the validation set for which the $D_{0.1cc}$ was dose limiting the method predicted this correctly. The same variation in plan quality observed in the training set likely occurred also for the patients in the validation set. Therefore if the predicted MLD or $D_{0.1cc}$ was lower than the achieved MLD (i.e. outside the confidence interval), this does not necessarily indicate that the method is over-optimistic, for the achieved plan may have been a low quality treatment plan. Therefore the achieved mean PTV doses were per definition not for 100% of the patients inside the confidence interval of the prediction. Ideally to test a plan quality prediction method, a set of treatment plans of prior patients should be available which are all Pareto optimal and for which the different planning objectives were prioritized consequently [14]. Unfortunately such a set is currently not available. As a result, errors between the expected value and achievable values may be caused by non-optimality of the validation data or by limitations of the model.

The DVH prediction model assumes that the dose–distance relation between an OAR and the PTV is independent on the patient specific anatomy. As long as the dependence on anatomy is absent or weak, the model could be used for any treatment site, OAR and treatment modality. In fact similar models have been presented for head-and-neck, prostate and pancreatic cancer patients [2,4,5]. However, a number of factors could potentially influence the dose–distance relation and therefore lead to prediction errors if not taken into account. First, the potential to spare an OAR depends also on the dose constraints and priorities of other OARs. If these priorities are consistent in the training database, the effect on the prediction will be small. However, if the priority of sparing different OARs differed per patient in the training database (with sacrificing one OAR to maximally spare another as ultimate example), the prediction accuracy of the model will decrease. Second, the location of the PTV in the lungs could affect the dose–distance relation. As shown in Supplementary material B sparing the lung tends to be easier for PTVs located more cranially than caudally ($r = 0.48$ to $-0.24$). No effect of the radial position of the PTV or the lung volume was observed. Third the dose–distance relation could also depend on the beam configuration and was therefore taken into account in some [8], but not in all prior studies [4]. For the patients in the current study there were weak and opposite ($r = -0.25$ to 0.4) effects of the total arc length and the variation of gantry angles with dose-distance metrics (Supplementary material B). In this study, these effects of other OARs, tumour location and beam configuration were ignored. Still the prediction accuracy was sufficient for use in clinical practice. Taking into account also other OARs, tumour location and beam configuration, will likely improve the prediction accuracy of the model, but at the expense of a more complicated model.

The trained prediction model as presented here could be directly used also by other centres, but only if the plans have comparable population dose-distance relations as the patients in the current study. Whether this is the case may depend on if the same TPS and beam optimization, treatment technique, and the allowed PTV coverage and PTV maximum dose constraints are used. If not, it is recommended to train the model with previously treated
patients from the centre where the model is to be used of from another centre with comparable treatment protocols and TPS. A training database of around 25 patients should be sufficient for representative predictions (see Supplementary material C). The model can be scaled for different prescription doses and since it only depends on the dose-distance relation, also be used by centres with different contouring strategies.

An alternative approach that may improve the quality of treatment plans is to use fully automated treatment planning [13,15,16]. These methods could yield treatment plans of consistent high quality with minimal user interaction, however these methods are currently only available in a limited number of institutes. Also for fully automated treatment planning, QA approaches as described in this study may be very useful e.g. to detect outliers or bugs in the software that may lead to systematic errors among multiple treatment plans.

In this study the median dose–distance relation was used to validate the model. For use in clinical practice it could be advisable to use a lower percentile e.g. 25% or 10% to aim for treatment plans that correspond to the best 25% or 10% previously treated patients. This would ensure over time an increase in treatment plan quality and a smaller variation in plan quality over the population.

Conclusion

A method was presented and validated that predicts OAR doses and the maximum achievable dose escalation (up to 69 Gy) for a large cohort advanced stage NSCLC patients treated with volumetric arc therapy. The MLD could be predicted with an accuracy of $-0.07 \pm 1.9$ Gy (mean ± SD) and the maximum dose escalation was accurate within one 2 Gy fraction for 79% of the patients. The method provides confidence intervals that allow ranking of a planned DVH compared to the DVHs plans of prior patients, independent on differences in tumour size, geometry and location.

Accurate prediction of achievable tumour dose escalation and OAR sparing can be valuable for treatment planning QA and for a variety of personalized strategies early in the treatment process such as: identification of patients that may benefit from adaptive RT and shared decision making based on expected side effects vs. the probability of cure.

Conflict of interest statement

Authors acknowledge financial support from the Dutch Cancer Society (KWF MAC 2011-4970 awarded to WvE and KWF EMCR 2014-6667 awarded to SP). Erasmus MC Cancer Institute also has research collaborations with Elekta and Accuray.

Acknowledgements

Authors acknowledge financial support from the Dutch Cancer Society (KWF MAC 2011-4970 awarded to WvE and KWF EMCR 2014-6667 awarded to SP). Erasmus MC Cancer Institute also has research collaborations with Elekta and Accuray.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2015.07.040.

References
