Toward personalized dose-prescription in locally advanced non-small cell lung cancer: Validation of published normal tissue complication probability models

M Thor[1], **J.O Deasy**[1], **A Iyer**[1], **E Bendau**[1], **A Fontanella**[1], **A Apte**[1], **E Yorke**[1], **A Rimner**[2], **A Jackson**[1]

[1]Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, US

[2]Dept. of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, US

Abstract

Purpose: To identify published normal tissue complication probability (NTCP) models suitable for patient-specific dose-prescription in locally advanced non-small cell lung cancer (LA-NSCLC) through in-house validation.

Material and Methods: From eight previously published candidate NTCP models (grade 2 acute esophagitis and radiation pneumonitis; AE2, RP2), patient-specific dose-responses were calculated using model variables and fractionation-corrected doses for 241 LA-NSCLC patients treated with chemo-IMRT to 50-80Gy@1.8-2.0Gy between 2004 and 2014 (AE2/RP2 rate: 50%/12%). A model was judged final if it significantly predicted AE2 or RP2 (p 0.05), was discriminative and well calibrated (AUC>0.60; Hosmer-Lemeshow test p_{HL}>0.05), which were assessed as the median over 1000 bootstrap samples.

Results: Models for AE2 had superior discrimination to RP2 models (AUC=0.63–0.65 vs. 0.51–0.65). The final AE2 model included mean esophageal dose and concurrent chemotherapy (AUC=0.65; p<0.0001). The final RP2 model was a slightly adjusted version of the RP2 model with the best discrimination, and included age, mean lung dose, and pulmonary comorbidity (AUC=0.73; p<0.0001).

Conclusion: Of the eight investigated and published NTCP models, one model successfully described AE2 and one slightly adjusted model successfully described RP2 in the independent cohort. Estimates from these two NTCP models will, therefore, be considered internally when prescribing patient-specific doses in LA-NSCLC patients.

Corresponding author: Maria Thor, Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, NYC, NY 10012, (+1) 646-888-8013, thorm@mskcc.org.

Introduction

Outcomes after radiotherapy (RT) for inoperable locally advanced non-small cell lung cancer (LA-NSCLC) have been poor, with associated median survival times barely approaching 30 months since randomization [1]. Efforts to further increase the efficacy of RT for these patients have focused on population-based dose-escalation, *e.g.*, RTOG's 0617 randomized controlled phase III trial in which a standard 60Gy was compared to an escalated 74Gy prescription dose [1]. However, given shorter than anticipated survival times in the 74Gy arm, a 60Gy prescription dose is still the standard almost 40 years after first being established [2].

A likely normal tissue-related explanation for the failure of dose-escalation in RTOG 0617 was given in their multivariate overall survival model: both heart dose and severe acute esophagitis (AE) lead to shorter survival times [1]. These results suggest that taking into account an individual's likelihood of developing normal tissue injuries may aid in better tailoring patient-specific prescription dose levels. Also, a secondary analysis of the RTOG 0617 data demonstrated that patients treated with intensity-modulated RT (IMRT) had significantly lower heart doses compared to patients treated with three-dimensional conformal RT (3DCRT) [3]. Therefore, the use of IMRT over 3DCRT is likely to further minimize normal tissue doses.

Patient-specific dose-prescription aims to tailor the total dose and the number of fractions to give the maximum achievable biologically equivalent tumor dose that respects current dose-volume constraints and patient-specific normal tissue complication probability model (NTCP) limits. Others have previously adopted such a 'model-based approach', based on multivariate models, for the purpose of selecting patients for photon or proton therapy [4], and there are also examples in which patient-specific dose levels have been suggested based on NTCP estimates from the Lyman-Kutcher-Burman model [5, 6]. To the best of our knowledge, incorporation of patient-specific NTCP estimates from multivariate models (including disease, patient and treatment characteristics) has, thus far, not been explored in order to estimate patient-specific dose-prescriptions.

The purpose of this study was to identify published NTCP models applicable to RT for lung cancer, and to validate these in a large, independent, inoperable LA-NSCLC cohort. Models with satisfactory performance will be incorporated and their associated estimates used in a planned prospective clinical protocol investigating patient-specific dose-prescription for this group of patients.

Methods and materials

The goal was to identify NTCP models that could be used to guide patient-specific dose-prescription in chemo-IMRT. A total of eight published NTCP models relevant to toxicities after LA-NSCLC RT were identified. Of these, four models attempted to predict grade 2 acute esophagitis (Models: $AE2_{M1-M4}$) [7–10], and four models grade 2 radiation pneumonitis (Models: $RP2_{M1-M4}$) [10–13].

Published acute esophagitis models

$AE2_{M1-M4}$ were logistic regression-based of type $1/(1+e^{-x})$; x in each of the four models was:

$$AE2_{M1}: -3.1 + (ConChemo \bullet 1.50) + (Dmean_{Eso} \bullet 0.07) \quad [7]$$

$$AE2_{M2}: -1.8 + (ConChemo \bullet 1.49) + (Dmean_{Eso} \bullet 0.09) \quad [8]$$

$$AE2_{M3}: -6.4 + (ConChemo \bullet 2.65) + (Dmean_{Eso} \bullet 0.12) + (Female \bullet 1.20) + \quad [9]$$
$$(Primary\ Tstage \geq 3 \bullet 0.99)$$

$$AE2_{M4}: -3.9 + (TxDays \bullet 0.05) + (D25_{Eso} \bullet 0.01) + (D40_{Eso} \bullet 0.04) \quad [10]$$

More specifically, AE_{M1-3} had been derived from three cohorts consisting of 374, 115, and 149 lung cancer patients (stage III NSCLC: 71%, 75% (stage II), 99%) treated with conventionally fractionated 3DCRT [7, 8] or IMRT [9]. The AE2 rates (corresponding to AE requiring medical attention or care) in these three cohorts were 32%, 82% and 36%. Concurrent chemotherapy (ConChemo) as opposed to no/sequential chemotherapy and mean esophageal dose ($Dmean_{Eso}$) predisposed for AE2 [7–9]. In addition, $AE2_{M3}$ encompassed a gender covariate and a primary tumor stage (T stage) covariate in which being female and having T stage 3, respectively, increased the risk of AE2 [9]. $AE2_{M4}$ included the minimum dose to the hottest 25% and 40% of the esophagus volume ($D25_{Eso}$, $D40_{Eso}$) and the number of treatment days (TxDays) [10]. This model had been derived from 278 patients (NSCLC: 78%) treated either with conventionally fractionated 3DCRT or IMRT, or with stereotactic body RT, and the associated AE2 rate for the same definition as in the three aforementioned studies was 22%.

Published radiation pneumonitis models

Similar to $AE2_{M1-M4}$, $RP2_{M1-M4}$ [10–13] were also logistic regression-based, and $RP2_{M1}$ and $RP2_{M2}$ [10, 11] were of type $1/(1+e^{-x})$ in which x was:

$$RP2_{M1}: -3.5 + (D10_{Heart} \bullet 0.02) + (Dmean_{Lung} \bullet 0.06) \quad [11]$$

$$RP2_{M2}: -3.8 + (Dmax_{Heart} \bullet 0.02) + (D65_{Lung} \bullet 0.3) + (TxDays \bullet 0.02) \quad [10]$$

RP2$_{M3}$ and RP2$_{M4}$ were also of type $1/(1+e^{-x})$, but were presented somewhat differently [12, 13]. More specifically, x in RP2$_{M3}$ was:

$$-3.9 + (\text{Dmean}_{\text{Lung}} \bullet 0.06)$$

This expression resulted in a dose for a 50% complication rate (D$_{50}$) of 30.8Gy with a steepness (γ_{50}) of 0.97 [12]. RP2$_{M4}$ was a direct development of RP2$_{M3}$ but the dose-response curve for RP2$_{M3}$ was modified to also incorporate effects of six disease, patient, and treatment characteristics represented as Odds Ratios (ORs) [13], adopted from a preceding meta-analysis [14], which involved approximately 1000 patients. The exact number of patients for each characteristic is given in Table 1. These characteristics were age (cut-point: 63 years; OR=1.66), current smoker (OR=0.62), former smoker (OR=0.69), inferior/middle tumors (OR=1.87), pre-existing pulmonary comorbidity (chronic obstructive pulmonary disease or general pre-existing lung disease: OR=2.27), and sequential as opposed to ConChemo (OR=1.60). All ORs were included multiplicatively; patient-specific dose responses were modified as follows:

$$D_{50}{}^{\text{risk}} = D_{50}{}^{0}(1 - 0.25/\gamma 50(\ln(\text{OR})))$$

$$\gamma_{50}{}^{\text{risk}} = \gamma_{50}{}^{0} - 0.25(\ln(\text{OR}))$$

Where D$_{50}{}^{0}$ and $\gamma_{50}{}^{0}$ were the values appropriate for patients with no risk factors. To reproduce QUANTEC's population based dose-response curve [12], Appelt *et al* [13] found that D$_{50}{}^{0}$=34.4Gy and $\gamma_{50}{}^{0}$=1.19.

For RP2$_{M1}$, the RP2 rate was 23% six months after primarily conventionally fractionated RT in 209 NSCLC patients, and RP2 had been aggregated from two scoring systems (requiring steroids/persisting cough requiring narcotic or dyspnea during minimal effort) [11]. The RP2 rate in RP2$_{M2}$ was 13%, which had been assessed as limiting activities of daily living and indicating medical intervention at a median of 4.7 months. The cohort is the same as the AE2$_{M4}$ model [10] and is described in more detail above. The data for which RP2$_{M3}$ was derived were synthesized from ten studies in which a total 1175 patients were most commonly treated with conventionally fractionated 3DCRT with an across study median RP2 rate of 16% [11]. Of note, a variety of scoring systems and grades had been used. For example, in the three cohorts that included >200 patients, RP was scored using two different scoring systems, a total of three grades, and three follow-up time intervals (no limit, six months, or one year). Notwithstanding, the phrasing RP2 will henceforth be used when referring to RP in all four models.

Independent validation cohort data

All stage IIIA and IIIB LA-NSCLC patients treated at the Memorial Sloan Kettering Cancer Center between 2004 and 2014 initially qualified for inclusion in this study [15, 16]. Since the ultimate goal was to guide patient-specific dose-prescription prospectively based on

NTCP models, inclusion was limited to patients treated with regimens likely to be used in the future, *i.e.*, chemotherapy combined with CT-based IMRT planning. A total of 241 patients fulfilled these criteria. These patients had been treated with chemo-IMRT (concurrent/sequential: 64%/36%) to a median of 64Gy (range: 50–80Gy) in 1.8Gy or 2.0Gy fractions. Dose was prescribed to the primary tumor and positron emission tomography (PET) positive lymph nodes. All dose distributions were re-calculated with our currently used dose calculation algorithm (*Eclipse AAA v.13, Varian Medical Systems, Palo Alto, CA, US*).

The esophagus was re-segmented to extend from the cricoid cartilage to the gastro-intestinal **junction**, the lung volume was defined as the total lung volume minus the gross target volume and tumor-involved lymph nodes, and the heart included all chambers and extended superiorly from the pulmonary artery split to its inferior terminus. Both AE2 and RP2 had been scored according to CTCAE v.4.03, and corresponded to altered eating/swallowing indicating oral supplements for AE, and limiting activities of daily living and indicating medical intervention for RP. The AE2 and RP2 rates were 50% and 12%, respectively. The distribution of model variables is summarized with those in the respective derivation cohorts in Table 1.

Evaluation of published models

Before evaluating the performance of each model, relevant dose-volume metrics in the independent cohort were converted from physical dose to equivalent doses in 2 Gy fractions assuming $\alpha/\beta=10$ Gy for AE2 [17], and $\alpha/\beta=3$ Gy for RP2 [12]. Evaluation of model performance adhered to the Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [18], and more specifically to a TRIPOD Type 4 model. In particular, calibration and discrimination were assessed. Calibration was evaluated in quintiles by the Hosmer-Lemeshow test (p_{HL}) [19], and discrimination by the area under the receiver-operating curve (AUC) and p-values from the correlation between observations and predictions. Significantly predictive models (p $\le$ 0.013 given four models/tests per endpoint) being well discriminative (AUC>0.60) and with $p_{HL}>0.05$ were considered final. If multiple models presented with the same AUC, the model better adhering to the identity line, and as encouraged by TRIPOD, in the plotted calibration curves between model predictions and observations was considered final. Published model coefficients were applied to the corresponding variables in the independent cohort, and internal generalizability was evaluated over 1000 bootstrap samples (*i.e.*, calibration and discrimination measures are reported as the median values across these 1000 samples). It should be emphasized that to enable fair comparisons of the performance across models, their associated coefficients were not re-fitted.

For the final models, re-fitting was ultimately performed and the resulting performance was compared to that of the original, *i.e.*, non re-fitted models. If not explicitly addressed by the final model, the following disease, patient, and treatment characteristics were additionally investigated (with the exception of Appelt *et al* [13], given that the model variables were assessed from a larger meta-analysis [14]): age, current smoker, former smoker, gender, histology, hypertension, Karnofksy performance status (KPS), tumor inferior-superior

location, tumor laterality, and tumor stage. The significance level was adjusted from 5% to take into account the number of tests performed for each model. This approach also used logistic regression with bootstrap resampling. All analyses were performed in MATLAB R2016a, and extraction of dose-volume data was performed in the computational environement for radiological research, CERR [20].

Results

Concurrent chemotherapy and mean esophageal dose predict acute esophagitis

All four AE2 models significantly explained AE2 in the independent cohort (p<0.0002; Table 2; Fig 1; Fig S1), and discrimination was similar across all models (AUC=0.63–0.65). From the calibration curves in Fig S1, predictions by $AE2_{M1}$ and $AE2_{M3}$ typically underestimated the observed AE2 rates for primarily low to intermediate risks, but $AE2_{M1}$ was overall well calibrated whereas $AE2_{M3}$ was not. In contrast, predictions by $AE2_{M2}$ and in particular $AE2_{M4}$ overestimated the observed AE2 rates. Since $AE2_{M1}$ presented with the overall highest discrimination and was well calibrated (AUC=0.65) this model was considered the final NTCP model for AE2.

Re-fitting $AE2_{M1}$ in the independent cohort generated somewhat different regression coefficients for both ConChemo and $Dmean_{ESO}$, but these differences did not produce substantially better model performance compared to the original $AE2_{M1}$ model ($AE2_{M1}$ *vs.* re-fitted: AUC=0.65 *vs.* 0.65). Also, replacing $Dmean_{ESO}$ with our best dose-volume metric $D40_{Eso}$, and refitting $D40_{Eso}$ together with ConChemo in the independent cohort did not improve model performance considerably (AUC=0.66). None of the nine additionally investigated disease/patient/treatment characteristics significantly predicted AE2.

Models combining dose with disease and patient characteristics better predict radiation pneumonitis relative to 'dose-only models'

Overall, discrimination of $RP2_{M1-M4}$ was lower compared to that of $AE2_{M1-M4}$ (AUC=0.51–0.65 vs. AUC=0.63–0.65; Table 2; Figs 1, 2, S1). Predictions by $RP2_{M2}$ and $RP2_{M3}$ overestimated and underestimated the observed RP2 rates, respectively (AUC=0.51 *vs*. 0.57). Although $RP2_{M1}$ presented with an AUC of only 0.51, this model was reasonably calibrated. Of all four RP2 models, $RP2_{M4}$ was the only discriminative model (AUC=0.65) and was at this stage deemed the final NTCP model for RP2.

However, since the $RP2_{M4}$ predictions typically underestimated the observed RP2 rates, each of the six model variables in $RP2_{M4}$ was further investigated. Only age and pulmonary comorbidity significantly predicted the observed RP2 rates (p=0.001, 0.003). For this analysis, age was (similar to the approach taken by Appelt *et al* [13]) dichotomized at 63 years, and the significance level was adjusted from 5% to 0.8% given six investigated variables/performed tests. The procedure performed in [13] was repeated but including only age and pulmonary comorbidity. The goal was to estimate the combination of $D_{50}{}^{0}$ and $\gamma_{50}{}^{0}$ (given only two risk factors in comparison to six for which $D_{50}{}^{0}$ and $\gamma_{50}{}^{0}$ were estimated in [13]) that would lead to a predicted RP2 risk as close as possible to that of QUANTEC's, and use these to assess patient-specific predicted RP2 risks. A more detailed explanation to

this procedure is given in the Supplementary material. The variable-specific prevalence and ORs were taken from [13]. Using instead these patient-specific RP2 risks, forming a modified RP2$_{M4}$ (RP2$_{M4mod}$), resulted in a significant relationship (p=0.0001) with superior discrimination and calibration compared to that of RP2$_{M4}$ (AUC=0.73 *vs.* 0.65; Fig 3). Hence, RP2$_{M4mod}$ was judged the final NTCP model for RP2.

Discussion

Previous attempts devoted to improve outcomes after RT for LA-NSCLC have mainly focused on population-based dose-escalation. However, such approaches have not proven viable [1]. Aiming instead at patient-specific dose-prescription, we in this study sought to identify relevant published NTCP models for use in determining dose prescriptions that would respect individual patients' tolerances based on NTCP estimates. Of eight identified NTCP models, one model focusing on acute esophagitis and one modified model focusing on radiation pneumonitis (AE2, RP2) successfully described AE2 and RP2 in the independent validation cohort, and were, therefore, considered final. Consequently, these two models will be considered suitable for use in the planning of a prospective protocol of patient-specific dose-prescription of LA-NSCLC patients.

The final AE2 model (AE2$_{M1}$) indicated that patients with higher Dmean$_{Eso}$ who are administered concurrent chemotherapy are more likely to develop AE2 [6]. Assuming an AE rate corresponding to that observed in the current cohort (AE2=50%), the goal according to AE2$_{M1}$ should be to limit Dmean$_{Eso}$ to below 23Gy for patients receiving concurrent chemotherapy. The three other investigated AE2 models [8–10] typically included concurrent chemotherapy with the same or other esophageal dose parameterizations, but these models did not describe the AE2 rates in our validation cohort as well as AE2$_{M1}$. Further, neither re-fitting AE2$_{M1}$, nor replacing Dmean$_{Eso}$ with the best univariate dose-volume predictor (D40$_{Eso}$) improved the performance over that of the original AE2$_{M1}$. None of the additionally investigated disease, patient, or treatment characteristics (age, current smoker, former smoker, gender, histology, hypertension, KPS, tumor inferior-superior location, tumor laterality, and tumor stage) significantly predicted AE2. Interestingly, even though the definition of AE2 differed to a lesser extent across the four investigated AE2 models compared to that of the four RP2 models, the performance of the final AE2 model was less satisfying compared to the final RP2 model (AUC=0.65 *vs.* 0.73). It is, therefore, likely that incorporation of factors currently unaccounted for may improve model performance. One source of additional and potentially prognostic information for this purpose may come from PET imaging of the esophagus. For instance, within a LA-NSCLC cohort of 79 patients, Niedzielski *et al* [21] found that mid-treatment PET distributions significantly predicted AE2 with AUCs in the range 0.91–0.92.

Of the four investigated RP2 models [10–13], the only model describing RP2 reasonably in the independent cohort was the RP2$_{M4}$ model [13]. In addition to Dmean$_{Lung}$, RP2$_{M4}$ also indicated that advanced age, having no smoking history, pre-existing pulmonary comorbidity, presenting with inferior or middle tumors as opposed to superior tumors, and being treated with sequential compared to concurrent chemotherapy, predisposed patients for RP2. Univariate analysis of each of these characteristics, however, indicated that only

advanced age and having pre-existing pulmonary comorbidity significantly predicted RP2 in our validation cohort. Our modified response, including only these two risk factors (as opposed to all six from [13]) resulted in both better discrimination and calibration (AUC=0.73 *vs.* 0.65), which lead to the selection of RP2$_{M4mod}$ as the best model for RP. Although model performance of this final RP2 model was superior to that of the final AE2 model (AUC=0.73 *vs.* 0.65), incorporation of other factors (such as pre-treatment imaging of the normal lung) may provide additional prognostic information for RP2. Similar to AE2, PET information (pre-treatment of the lung) has been found to predict mild dyspnea [22]. In addition, interstitial lung disease assessed from pre-treatment CT scans has been associated with severe and fatal RP [23]. If any image-derived information (or any other information not investigated here) proves predictive, it could be incorporated into RP2$_{M4mod}$ in a similar way, as this model was developed from RP2$_{M4}$ in this study (Supplementary material).

Conclusions

In the effort to identify published NTCP models appropriate for patient-specific dose-prescription respecting individual patient's tolerance doses in RT for lung cancer, one AE2 model and one modified RP2 model were found to be relevant. Similar external efforts are encouraged, but model performance should be re-assessed particularly in cohorts considerably different from the LA-NSCLC cohort investigated here. As a starting point, to manage AE2, a lower mean esophageal dose will be considered for patients assigned to concurrent chemotherapy. The modified RP2 model indicated that in order to manage the rate of RP2, the mean lung dose should be further minimized in particular for older patients presenting with pulmonary comorbidity. Our overall prospective goal will be to continuously collect new patient-specific prognostic information that will improve the performance of models, and use them to improve the treatment and accordingly the outcomes for LA-NSCLC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Bradley JD, Paulus R, Komaki R, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small cell lung cancer (RTOG 0617): a randomized, two-by-two factorial phase 3 study. Lancet Oncol 2015;16:187–99 [PubMed: 25601342]

[2]. Perez CA, Stanley K, Rubin P, et al. A prospective randomized study of various irradiation doses and fractionation schedules in the treatment of inoperable non-oat-cell carcinoma of the lung. Preliminary report by the Radiation Therapy Oncology Group. Cancer 1980; 5:2744–53

[3]. Chun SG, Hu C, Choy H, et al. Impact of intensity-modulated radiation therapy technique for locally advanced non-small-cell lung cancer: A secondary analysis of the NRG Oncology RTOG 0617 randomized clinical trial. J Clin Oncol 2017;35:55–62

[4]. Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, and Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013; 107:267–73 [PubMed: 23759662]

[5]. Ten Haken RK, Martel MK, Kessler ML, et al. Use of Veff and iso-NTCP in the implementation of dose escalation protocols. Int J Radiat Oncol Biol Phys 1993;27:689–95 [PubMed: 8226166]

[6]. Rosenzweig KE, Mychalczak B, Fuks Z, et al. Final report of the 70.2-Gy and 75.6-Gy dose levels of a phase I dose escalation study using three-dimensional conformal radiotherapy in the treatment of inoperable non-small cell lung cancer. Cancer J 2000;6:82–7 [PubMed: 11069224]

[7]. Huang EX, Bradley JD, El Naqa I, et al. Modeling the risk of radiation-induced acute esophagitis for combined Washington University and RTOG trial 93–11 lung cancer patients. Int J Radiat Oncol Biol Phys 2012;82:1674–9. [PubMed: 21658856]

[8]. Huang EX, Robinson CG, Molotievschi A, et al. Independent test of a model to predict severe acute esophagitis. Adv Radiat Oncol 2017;2:37–43 [PubMed: 28740914]

[9]. Wijsman R, Dankers F, Troost EG, et al. Multivariable normal-tissue complication modeling of acute esophageal toxicity in advanced stage non-small cell lung cancer patients treated with intensity-modulated (chemo-)radiotherapy. Radiother Oncol 2015;117:49–54 [PubMed: 26341608]

[10]. von Reibnitz D, Yorke ED, Oh JH, et al. Serum Alpha-2-Macroglobulin as an intrinsic radioprotective factor in patients undergoing thoracic radiation therapy. BioRxiv 656090 [Preprint]. 5 31, 2019 Available from: 10.1101/656090

[11]. Huang EX, Hope AJ, Lindsay PE, et al. Heart irradiation as a risk factor for radiation pneumonitis. Acta Oncol 2011;50:51–60. [PubMed: 20874426]

[12]. Marks LB, Bentzen SM, Deasy JO, et al. Radiation dose-volume effects in the lung. Int J Radiat Oncol Biol Phys 2010;76(3 Suppl):70–6.

[13]. Appelt AL, Vogelius IR, Farr KO, Khalil AA, and Bentzen SM. Towards individualized dose constraints: Adjusting the QUANTEC radiation pneumonitis model for clinical risk factors. Acta Oncol 2014;53:605–12. [PubMed: 23957623]

[14]. Vogelius IR, and Bentzen SM. A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis. Acta Oncol 2012;51:975–83. [PubMed: 22950387]

[15]. Oberije C, De Ruysscher D, Houben R, et al. A validated prediction model for overall survival from stage III non-small cell lung cancer: Toward survival prediction for individual patients. Int J Radiat Oncol Biol Phys 2015;92:935–44. [PubMed: 25936599]

[16]. Sonnick MA, Oro F, Yan B, et al. Identifying the optimal radiation dose in locally advanced non-small-cell lung cancer treated with definitive radiotherapy without concurrent chemotherapy. Clin Lung Cancer 2018; 7 6 [Epub ahead of print]

[17]. Werner-Wasik M, Yorke E, Deasy J, Nam J, and Marks LB. Radiation dose-volume effects in the esophagus. Int J Radiat Oncol Biol Phys 2010;76(3 Suppl):86–93

[18]. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162:1–73. [PubMed: 25560711]

[19]. Hosmer DW and Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. Commun in Stats 1980;10:1043–69.

[20]. Deasy JO, Blanco AI and Clark VH. CERR: a computational environement for radiotherapy research. Med Phys 2003;30: 979–85. [PubMed: 12773007]

[21]. Niedzielski JA, Yang J, Liao Z, et al. (18F)Fluorodeoxyglucose positron emission tomography can quantify and predict esophageal injury during radiation therapy. Int J Radiat Oncol Biol Phys 2016;96:670–80. [PubMed: 27681764]

[22]. Petit SF, van Elmpt WJ, Oberije CJ, et al. [18F]Fluorodeoxyglucose uptake patterns in lung before radiotherapy identify areas more susceptible to radiation-induced lung toxicity in non-small-cell lung cancer patients. Int J Radiat Oncol Biol Phys 2011;81:698–05. [PubMed: 20884128]

[23]. Glick D, Lyen S, Kandel S, et al. Impact of pretreatment interstitial lung disease on radiation pneumonitis and survival in patients treated with lung stereotactic body radiation therapy. Clin Lung Cancer 2018;e219–26 [PubMed: 29066051]
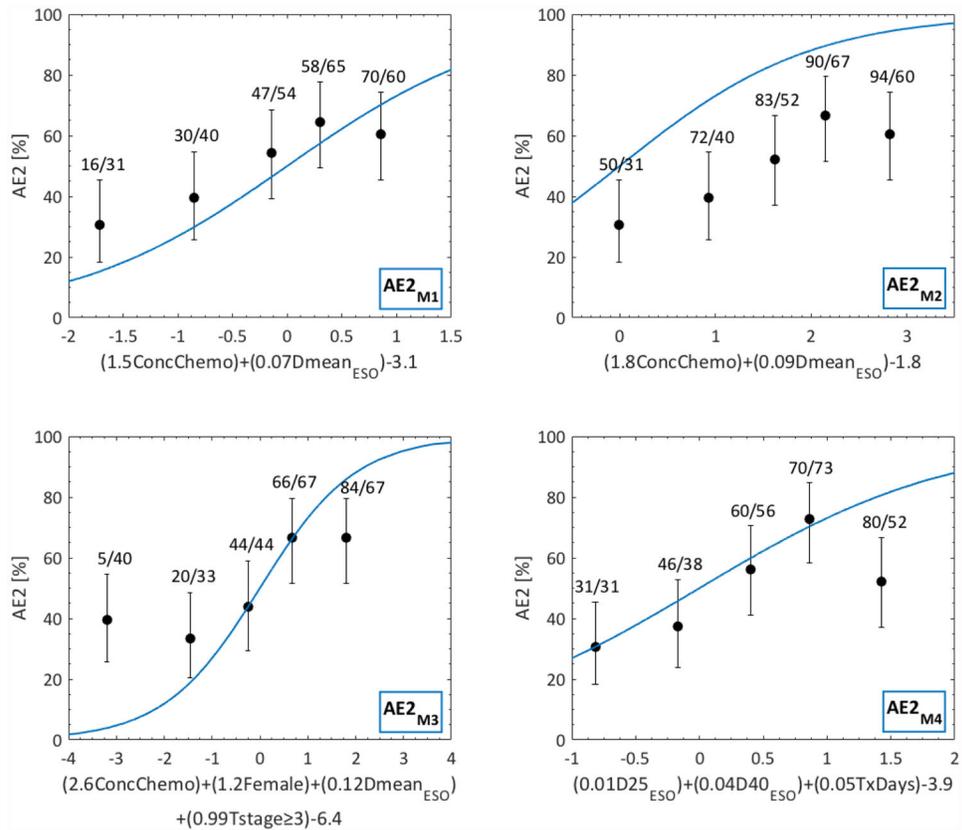
Fig 1.
Dose-response curves for the four published AE2 models (AE2$_{M1-M4}$) as applied to the independent cohort. Note: AE2 predictions are given by blue lines; observed AE2 is denoted by black quintiles (error bars: 95% binomial confidence intervals); quintile-specific predicted/observed AE2 rates [%] are given above each quintile; the argument of each model's logistic-regression based equation is given on the x-axis
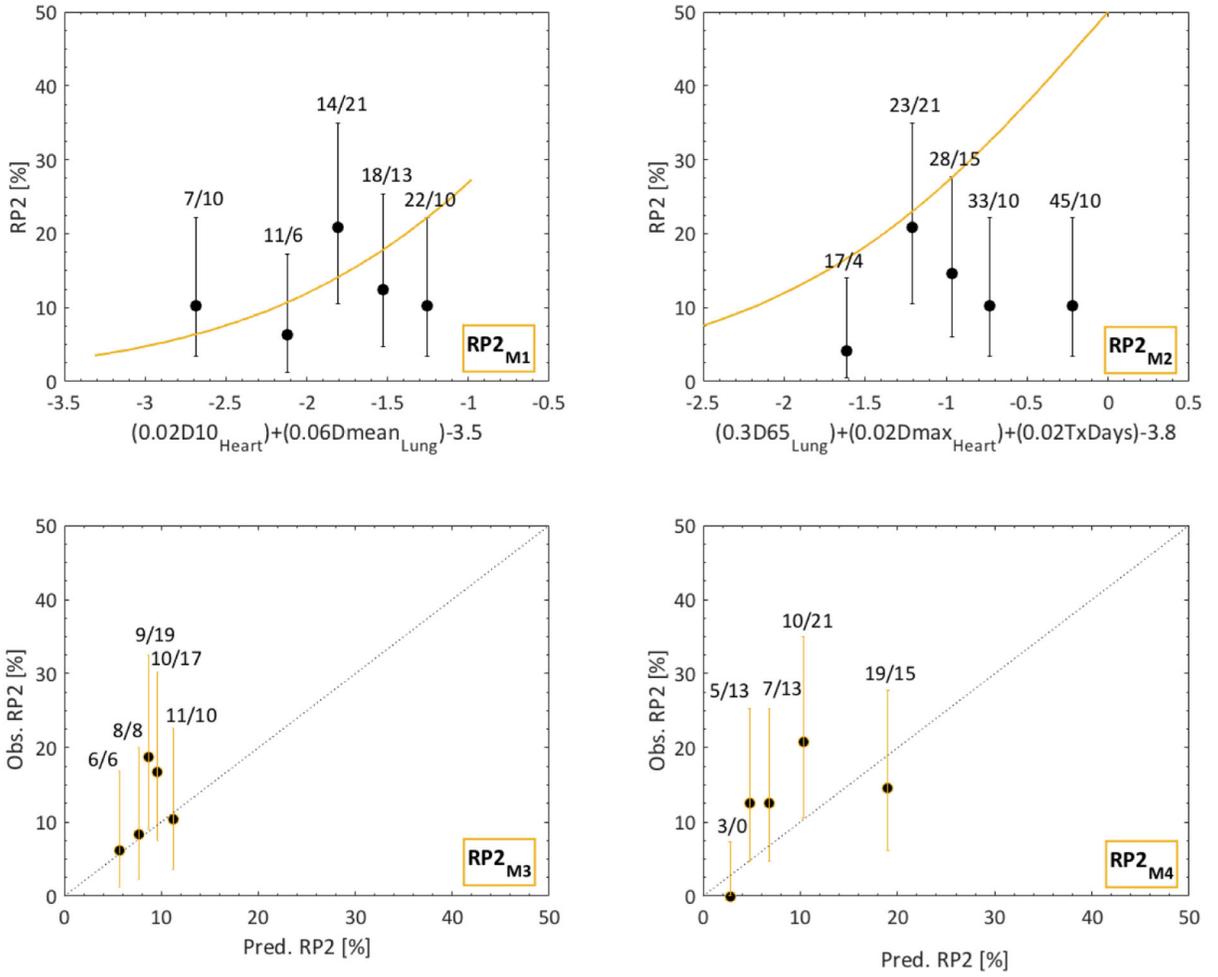
Fig 2.
Dose-response curves for the four published RP2 models ($RP2_{M1-M4}$) as applied to the independent cohort. Note: For $RP2_{M3, M4}$, calibration curves are instead given; RP2 predictions are given by orange lines; observed RP2 is denoted by black quintiles (error bars: 95% binomial confidence intervals); quintile-specific predicted/observed RP2 rates [%] are given above each quintile; black dotted lines for RP_{M3} and RP_{M4} are identity lines; for RP_{M1} and RP_{M2} the argument of each model's logistic-regression based equation is given on the x-axis
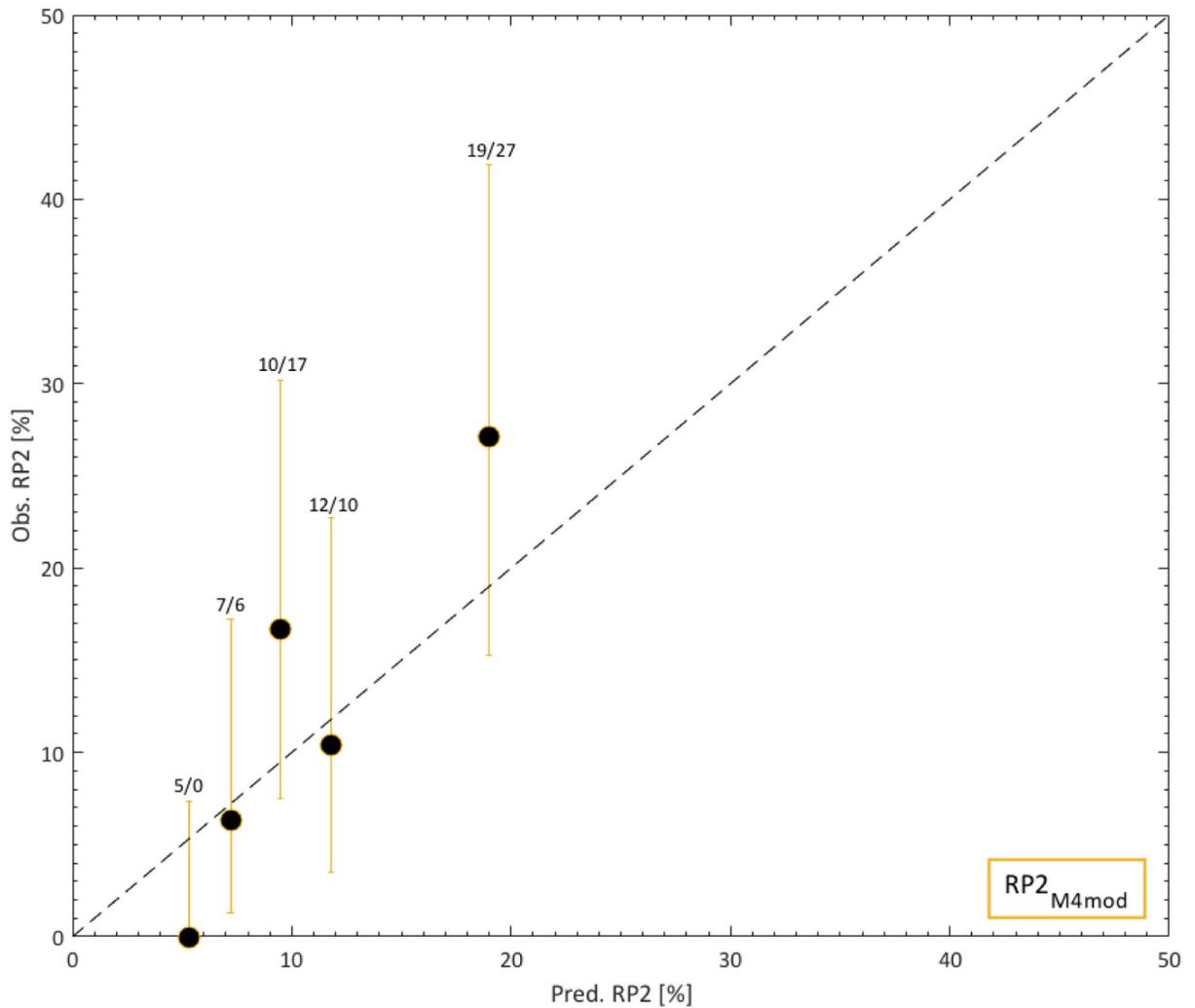
Fig 3.
Calibration curve for the modified $RP2_{M4}$, i.e., $RP2_{M4mod}$ based on the prevalence and ORs only for age and pulmonary comorbidity in the independent cohort. Note: RP2 predictions are given by orange lines; observed RP2 is denoted by black quintiles (error bars: 95% binomial confidence intervals); quintile-specific predicted/observed RP2 rates are given above each quintile; the black dotted line is an identity line.

Table 1.

Distribution of variables in the identified AE2 and RP2 models (AE2$_{M1-M4}$; RP2$_{M1-M4}$) in each derivation cohort, if available, and in the independent cohort.

Model [ref]	Model variables	Derivation cohort	Independent cohort
AE2$_{M1}$ [7]	1. Concurrent chemotherapy	24 (90)	64 (155)
	2. Dmean$_{ESO}$ [Gy]	25 (range: 0.02–62)	27 (8)
AE2$_{M2}$ [8]	1. Concurrent chemotherapy	74 (85)	64 (155)
	2. Dmean$_{ESO}$ [Gy]	22 (12)	27 (8)
AE2$_{M3}$ [9]	1. Concurrent chemotherapy	62 (92)	64 (155)
	2. Dmean$_{ESO}$ [Gy]	N/A	27 (8)
	3. Female	35 (52)	52 (125)
	4. Primary T stage	48 (72)	46 (111)
AE2$_{M4}$ [10]	1. Treatment days [d]	24 (17)	46 (6)
	2. D25$_{ESO}$ [Gy]	25 (22)	52 (13)
	3. D40$_{ESO}$ [Gy]	15 (18)	37 (19)
RP2$_{M1}$ [11]	1. D10$_{Heart}$ [Gy]	41 (range: 0.1–81)	36 (18)
	2. Dmean$_{Lung}$ [Gy]	median: 18 (range: 4–33)	12 (2)
RP2$_{M2}$ [10]	1. Dmax$_{Heart}$ [Gy]	48 (36)	65 (13)
	2. D65$_{Lung}$ [Gy]	2 (2)	3 (2)
	3. Treatment days [d]	24 (17)	46 (6)
RP2$_{M3}$ [12]	1. Dmean$_{Lung}$ [Gy]	N/A	12 (2)
RP2$_{M4}$ [13]	1. Dmean$_{Lung}$ [Gy]	N/A	12 (2)
	2. Age>63y	N/A; mean: 65y; 8 studies	62 (149); mean: 66y
	3. Current smoker	28 (150; 3 studies)	31 (74)
	4. Former smoker	66 (221; 2 studies)	62 (149)
	5. Inferior/middle tumor	44 (297; 4 studies)	35 (84)
	6. Pulmonary comorbidity	26 (120; 4 studies)	28 (68)[*]
	7. Sequential chemotherapy	26 (287; 8 studies)	36 (86)

Note: Distribution of continuous variables is given as the mean (standard deviation), and binary variables as % (n) unless otherwise specified; all dose-volume metrics are presented as EQD2 in the independent cohort (cf. Material and Methods for further details), and as reported (typically not fractionated corrected) for each model. For [13], the number of studies for which the prevalence was estimated is given as well.

[*] Pulmonary comorbidity: Asthma, bronchiectasis, or chronic obstructive pulmonary disease.

Table 2.

Validation of the eight published AE2 and RP2 models ($AE2_{M1-M4}$; $RP2_{M1-M4}$) in the independent cohort.

Model [ref]	Model variables	Independent cohort		
		AUC	p-value	p_{HL}
$AE2_{M1}$ [7]	1. Concurrent chemotherapy	0.65	$5.3e^{-5}$	0.60
	2. Dmean$_{ESO}$ [Gy]			
$AE2_{M2}$ [8]	1. Concurrent chemotherapy	0.64	$5.8e^{-5}$	0.29
	2. Dmean$_{ESO}$ [Gy]			
$AE2_{M3}$ [9]	1. Concurrent chemotherapy	0.65	$4.6e^{-5}$	0.99
	2. Dmean$_{ESO}$ [Gy]			
	3. Female			
	4. Primary T stage 3			
$AE2_{M4}$ [10]	1. Treatment days [d]	0.63	$1.8e^{-4}$	0.34
	2. D25$_{ESO}$ [Gy]			
	3. D40$_{ESO}$ [Gy]			
$RP2_{M1}$ [11]	1. D10$_{Heart}$ [Gy]	0.51	0.56	0.68
	2. Dmean$_{Lung}$ [Gy]			
$RP2_{M2}$ [10]	1. Dmax$_{Heart}$ [Gy]	0.51	0.60	0.37
	2. D65$_{Lung}$ [Gy]			
	3. Treatment days [d]			
$RP2_{M3}$ [12]	1. Dmean$_{Lung}$ [Gy]	0.57	0.39	0.89
$RP2_{M4}$ [13]	1. Dmean$_{Lung}$ [Gy]	0.65	0.07	0.90
	2. Age>63y			
	3. Current smoker			
	4. Former smoker			
	5. Inferior/middle tumor			
	6. Pulmonary comorbidity			
	7. Sequential chemotherapy			

Note: AUC, p-values (from the correlation between observations and predictions), and pHL are given as the median across the 1000 Bootstrap samples.