University of Groningen

External validation of NTCP-models for radiation pneumonitis in lung cancer patients treated with chemoradiotherapy

Niezink, Anne G H; van der Schaaf, Arjen; Wijsman, Robin; Chouvalova, Olga; van der Wekken, Anthonie J; Rutgers, Steven R; Pieterman, Remge M; van Putten, John W G; de Hosson, Sander M; van der Leest, Annija H D

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

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

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Niezink, A. G. H., van der Schaaf, A., Wijsman, R., Chouvalova, O., van der Wekken, A. J., Rutgers, S. R., Pieterman, R. M., van Putten, J. W. G., de Hosson, S. M., van der Leest, A. H. D., Ubbels, J. F., Woltman-van Iersel, M., Widder, J., Langendijk, J. A., & Muijs, C. T. (2023). External validation of NTCP-models for radiation pneumonitis in lung cancer patients treated with chemoradiotherapy. *Radiotherapy and Oncology*, *186*, Article 109735. Advance online publication. https://doi.org/10.1016/j.radonc.2023.109735

University of Groningen

External validation of NTCP-models for radiation pneumonitis in lung cancer patients treated with chemoradiotherapy

Niezink, Anne G H; van der Schaaf, Arjen; Wijsman, Robin; Chouvalova, Olga; van der Wekken, Anthonie J; Rutgers, Steven R; Pieterman, Remge M; van Putten, John W G; de Hosson, Sander M; van der Leest, Annija H D

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

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

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Niezink, A. G. H., van der Schaaf, A., Wijsman, R., Chouvalova, O., van der Wekken, A. J., Rutgers, S. R., Pieterman, R. M., van Putten, J. W. G., de Hosson, S. M., van der Leest, A. H. D., Ubbels, J. F., Woltman-van Iersel, M., Widder, J., Langendijk, J. A., & Muijs, C. T. (2023). External validation of NTCP-models for radiation pneumonitis in lung cancer patients treated with chemoradiotherapy. *Radiotherapy and Oncology*, *186*, Article 109735. Advance online publication. https://doi.org/10.1016/j.radonc.2023.109735

Original Article

External validation of NTCP-models for radiation pneumonitis in lung cancer patients treated with chemoradiotherapy

Anne G.H. Niezink [a,*], Arjen van der Schaaf [a], Robin Wijsman [a], Olga Chouvalova [a], Anthonie J. van der Wekken [b], Steven R. Rutgers [c], Remge M. Pieterman [d], John W.G. van Putten [e], Sander M. de Hosson [f], Annija H.D. van der Leest [a], Jan F. Ubbels [a], Marleen Woltman-van Iersel [a], Joachim Widder [a,g], Johannes A. Langendijk [a], Christina T Muijs [a]

[a] Department of Radiation Oncology; [b] Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen; [c] Department of Pulmonology, Treant Hospital Group, Scheper Hospital, Emmen; [d] Department of Pulmonary Diseases, Ommelander Hospital Groningen, Scheemda; [e] Department of Pulmonary Diseases, Martini Hospital Groningen, Groningen; [f] Department of Pulmonary Diseases, Wilhelmina Hospital Assen, Assen, the Netherlands; [g] Department of Radiation Oncology, Comprehensive Cancer Center Vienna, Medical University of Vienna, Austria

ARTICLE INFO

ABSTRACT

Purpose: Normal tissue complication probability (NTCP) models can be used to estimate the risk of radiation pneumonitis (RP). The aim of this study was to externally validate the most frequently used prediction models for RP, i.e., the QUANTEC and APPELT models, in a large cohort of lung cancer patients treated with IMRT or VMAT. [1–2]

Methods and materials: This prospective cohort study, included lung cancer patients treated between 2013 and 2018. A closed testing procedure was performed to test the need for model updating. To improve model performance, modification or removal of variables was considered. Performance measures included tests for goodness of fit, discrimination, and calibration.

Results: In this cohort of 612 patients, the incidence of RP $\geq$ grade 2 was 14.5%. For the QUANTEC-model, recalibration was recommended which resulted in a revised intercept and adjusted regression coefficient (from 0.126 to 0.224) of the mean lung dose (MLD),. The APPELT-model needed revision including model updating with modification and elimination of variables. After revision, the New RP-model included the following predictors (and regression coefficients): MLD (B = 0.250), age (B = 0.049, and smoking status (B = 0.902). The discrimination of the updated APPELT-model was higher compared to the recalibrated QUANTEC-model (AUC: 0.79 vs. 0.73).

Conclusions: This study demonstrated that both the QUANTEC- and APPELT-model needed revision. Next to changes of the intercept and regression coefficients, the APPELT model improved further by model updating and performed better than the recalibrated QUANTEC model. This New RP-model is widely applicable containing non-tumour site specific variables, which can easily be collected.

© 2023 The Author(s). Published by Elsevier B.V. Radiotherapy and Oncology 186 (2023) 1–8 This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Radiation pneumonitis (RP) is one of the most frequently reported radiation-induced toxicities after thoracic radiotherapy, depending predominantly on dose to the irradiated lungs [3]. To prevent RP, more advanced radiation techniques, such as Volumetric Modulated Arc Therapy (VMAT) and proton beam therapy (PBT), are increasingly applied to reduce the radiation dose to relevant normal tissues. Next to improvements in radiation techniques, knowledge of the most relevant dosimetric and clinical parameters is essential to guide radiotherapy treatment planning.

Modelling normal tissue complication probabilities (NTCP) renders potentially actionable parameters that can be used for treatment plan optimization which aims to minimise toxicity without jeopardizing locoregional tumour control. [4–5].

The seminal QUANTEC-project (QUantitative Analysis of Normal Tissue Effects in the Clinic) described an NTCP-model for "symptomatic RP" as depending solely on the mean lung radiation dose (MLD) [2]. Based on this model, a threshold for the MLD of 20 Gy is routinely used in daily clinical practice to limit the risk of symptomatic RP to less than 20%. [2] More recently, this QUANTEC-model was expanded by Appelt et al. using clinical factors retrieved from literature (APPELT-model).

Both, the QUANTEC- and APPELT-model were developed and validated in patient cohorts treated with conventional radiother-

* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, PO BOX 30001, 9300 RB Groningen, The Netherlands (A.G.H. Niezink).

E-mail address: a.g.h.niezink@umcg.nl (A.G.H. Niezink).

apy techniques (2D-RT or 3D-CRT) [6]. However, using these predictive models in clinical practice, requires confirmation of the generalizability of these models in patient cohorts treated with current standard modern radiation techniques such as Intensity Modulated Radiation Therapy (IMRT), VMAT, or PBT. [7–8].

Appelt et al. performed an external validation of their own model. However, the authors only had access to a relatively small retrospective dataset of patients treated with 3D-CRT (n = 103). [1] Thor et al. also performed an external validation, assessing discrimination and calibration of both the APPELT- and QUANTEC-model in a retrospective cohort of 241 patients treated with IMRT [9]. They concluded that both models underestimated the RP risk and that an adjusted APPELT-model performed best (AUC of 0.73). However, in this external validation, the need for model revision was not tested. [10].

Given these shortcomings, there is an unmet need for a more comprehensive external validation using prospectively collected data of patients treated with VMAT or IMRT. [6,11].

Therefore, the purpose of the current study was to externally validate and update the QUANTEC- and APPELT-model for RP in lung cancer patients treated with IMRT or VMAT using our prospective cohort of lung cancer patients included in the proPED lung program; NCT02421718. [12].

Material and methods

Patient selection

The study cohort was composed of patients included in a prospective data collection platform (proPED-lung). [12] Included were all patients treated with curative intent with (inoperable) stage I-IIIB and IV non-small-cell lung cancer (NSCLC) or small-cell lung cancer (SCLC). All patients were treated with definitive radiotherapy or chemo-radiotherapy between February 2013 and November 2018 with a minimal follow up of 6 months. [12] Patients treated with stereotactic body radiation therapy (SBRT), or 3D-CRT and patients treated with immunotherapy within 6 months after completion of radiotherapy were excluded.

The proPED-lung (NCT02421718) was reviewed by the hospital's institutional review board and was declared exempt from ethics committee approval. This program includes prospectively scored data on pre-defined time points on patient and tumour characteristics, treatment data (including radiotherapy dose-volume histogram (DVH) parameters and chemotherapy regimens), survival and toxicity. Since the introduction of the European General Data Protection Regulation in May 2018, written informed consent was obtained in all patients.

Treatment

All patients received high-dose radiotherapy, 45 to 60 Gy in 25 to 30 fractions with or without chemotherapy. In NSCLC patients, chemotherapy consisted of one or two cycles of platinum-based induction chemotherapy regimens, combined with gemcitabine, pemetrexed, or etoposide, followed by concurrent chemotherapy using weekly low dose gemcitabine, or cisplatin and docetaxel. Patients with SCLC were typically treated with four cycles of cisplatin and etoposide, starting radiotherapy at the second cycle.

All patients underwent a planning 4D computed tomography scan (4DCT) and a diagnostic [18]FDG-PET-CT was matched for delineation of the gross tumour volume (GTV), which encompassed the primary tumour and pathological lymph nodes. An internal target volume (ITV) was created based on the 10 breathing phases of the 4DCT. A 5- and 6-mm margin were added to create the clinical target volume (CTV) and the planning target volume (PTV), respec-

tively. The lungs were automatically contoured, excluding the gross tumour volume (Lungs-GTV).

Radiation pneumonitis

All patients were prospectively followed according to the Dutch guidelines for lung cancer by their referring pulmonologists. Follow-up included imaging with CT between 6 weeks to 4 months after the end of treatment and at least yearly thereafter. The reports and images of these visits were consecutively retrieved and systematically reviewed by three radiation oncologists (OC, RW and AN). RP was classified according to the CTCAE v4.0, from 6 weeks until 6 months after treatment; grade 0: no complaints; grade 1: complaints (i.e., increased shortness of breath, increased cough, fever) combined with changes on CT or X-ray images (CT and X-ray reports were checked and when deemed necessary the images were reviewed), not requiring steroids; grade 2: complaints and imaging changes requiring steroids; grade 3: requiring oxygen plus steroids. To avoid bias, RP was generally scored without reviewing treatment plans or dose volume histograms.

Statistical analysis

Two available NTCP-models for RP were externally validated. The first model was the QUANTEC pneumonitis model (QUANTEC-model) [13], which was a pooled analysis after refitting 10 different retrospective datasets. The MLD in Gy was included in the model as follows:

$$NTCP_{QUANTEC} = (1 + \exp(3.87 - 0.126 * MLD))^{-1}$$

The second model (APPELT-model) included variables based on a literature search. These variables included MLD (Gy), smoking status (current or former vs. never smoker), pulmonary comorbidity (any), age (>63 years), chemotherapy sequence (sequential as opposed to concomitant chemotherapy) and location of the tumour (mid or inferior vs. superior) [1]:

$$NTCP_{APPELT} = (1 + \exp(4.76 - 0.138 * MLD + 0.48$$
$$* current\ Smoker + 0.37 * former\ Smoker$$
$$- 0.82 * pulm\ CoMorb - 0.51 * old\ Age$$
$$- 0.47 * sequ\ Chemo - 0.63 * MidOrInf))^{-1}$$

Multivariate imputation by chained equations was performed 10 times for missing data to use all available patient data and to minimise the risk of bias, according to the method by van Buuren et al. [14–15]. All analyses were performed on all imputation sets and pooled following Rubin's rules [15–16].

A closed testing procedure, based on likelihood ratio tests, was performed to test the need for model updating, either by calibration-in-the-large (re-estimation of model intercept), recalibration (re-estimation of intercept and slope) or model revision (re-estimation of all coefficients) [10]. Based on the outcome of the closed testing procedure, model updates were performed accordingly.

Furthermore, modification or removal of variables to improve model performance was considered for model updates. First, manual stepwise variable elimination was used to remove non-significant variables. Optionally, categorical variables were regrouped, and the dichotomised variable age was replaced by a continuous variable. Considered transformation of the continuous variables included the square root, the second or third power, and the log of the different variables. To select the best performing variable, both univariable and multivariable analyses were performed and selection was based on Bayes' Information Criterion (BIC) with modified degrees of freedom (+2 for log, +1 for other transformations) as proposed by van den Bosch et al. [15].

Multiple performance measures of the original and updated models were calculated. Goodness of fit was tested using log-likelihood (LL), Akaike Information Criterion (AIC), and Bayesian information criterion (BIC). Discrimination was tested using the area under the receiver operating characteristic curve (AUC). Furthermore calibration-in-the-large (adjustment of the intercept only), calibration intercept and slope, and the Hosmer-Lemeshow (HL) test (with 10 risk groups) were used to assess model calibration. Nagelkerke's R^2 was calculated as pseudo measure of explained variance.

Shrinkage was applied to the final updated models using Ridge regression to compensate for optimism introduced by refitting. As an additional check for optimism internal validation using bootstrapping was performed.

For the statistical analyses SPSS version 23 and R (version 4.0.4) were used.

Results

At the time of analysis, 1,392 patients were included in the proPED-lung database, of which 1,322 had at least 6 months follow up. In total 691 patients were excluded because they did not meet the inclusion criteria (Fig. 1). Additionally, patients who did not start or stopped early during treatment or died during or within two weeks after treatment (n = 11, n = 2, and n = 6, respectively) were excluded, leaving 612 patients for the final evaluation.

The median age of the final study cohort was 68 years (range, 37–88 years). Most patients were male (59.0%), WHO performance was ≤ 1 in 84.6%, and most patients were former (64.5%) or current smokers (31.9%), at time of diagnosis (Table 1).

Chemo-radiotherapy was administered in 85.8% of the patients. Most patients (77.8%) were treated to a total dose of 60 Gy. During the study period, the treatment technique changed from a hybrid technique of 3DCRT combined with IMRT or VMAT (67.5%) to full IMRT or VMAT (32.5%). The median Mean Lung Dose (MLD) of the treatment plans was 11.5 Gy (SD: 3.8 Gy) and the median mean heart dose (MHD) was 6.1 Gy (SD: 6.7 Gy) (Table 1).

Grade 2 or higher RP was scored in 89 patients (14.5%), including 73 (11.9%) with grade 2, 11 patients (1.8%) with grade 3 and 2 patients (0.3%) with grade 4 RP. Three patients (0.5%) died from RP (grade 5).

In preparing an update of the models, variables were recoded into alternative representations. Age was used as a continuous variable. Smoking was regrouped in binary variables, (current smokers vs. others and current and recently stopped smokers (<3 months) vs. other, respectively). MLD and age were transformed as described in the Materials & Methods section. The results of the univariable analysis of all recoded or transformed parameters

Table 1
Patient, tumour and treatment characteristics.

Patient characteristics		All (n = 612)
Age (years)	Median (range)	68 (37–88)
	≥ 63	426 (69.6%)
Gender	Male	361 (59.0%)
WHO performance status	≤ 1	518 (84.6%)
Pulmonary comorbidity	Yes	219 (35.8%)
Cardiac comorbidity	Yes	275 (44.9%)
Smoking	Never	16 (2.6%)
	Former ≥ 3 months	289 (47.2%)
	Former < 3 months	106 (17.3%)
	Current	195 (31.9%)
	Unknown	6 (1.0%)
Tumour characteristics		
Location	Upper Lobe	335 (54.7%)
	Lower or middle lobe	167 (43.6%)
	Unknown/Tx	10 (1.6%)
Tumour type	NSCLC	521 (85.1%)
	SCLC	91 (14.9%)
Stage AJCC 7 or 8	I	26 (4.2%)
	II	94 (15.4%)
	III	429 (70.1%)
	IV*	63 (10.3%)
Treatment		
Technique	part IMRT/Part VMAT	413 (67.5%)
	full IMRT	94 (15.4%)
	full VMAT	105 (17.2%)
Dose	≤ 45	68 (11.1%)
	45–60	68 (11.1%)
	≥ 60	476 (77.8%)
Chemotherapy	Yes	525 (85.8%)
	Sequential	68 (13.0%)
	Concurrent	93 (17.7%)
	Induction and concurrent	364 (69.3%)
Lung dose volume parameters (median and SD)	Mean lung dose	11.5 (3.8)
	V5	46.8 (16.0)
	V20	19.3 (7.3)
	V30	14.2 (6.1)
	V40	9.3 (5.1)
Heart dose volume parameters (median and SD)	Mean heart dose	6.1 (6.7)
	V5	25.2 (27.2)
	V40	2.1 (6.7)

Abbreviations: NSCLC = non-small cell lung cancer, SCLC = small cell lung cancer, 3D-CRT = three dimensional conformal radiation therapy, IMRT = intensity modulated radiation therapy, VMAT = volumetric modulated arc therapy. * patients treated with synchronous metastases treated with curative intend like stage III.

Fig. 1. Flow chart of patient exclusion. SBRT = stereotactic body radiation therapy, 3D-CRT = three-dimensional conformal radiation therapy, RT = radiotherapy.

are shown in the supplementary Table 1. The best performing variables were selected for model update: the MLD without any transformation, age as a continuous predictor, and smoking divided in two categories (current smoker and quit < 3 months before treatment *versus* never-smoker and quit > 3 months).

For the QUANTEC-model, the closed testing procedure indicated that recalibration was required, including revision of the intercept and the regression coefficient of the MLD (Table 2a). Discrimination of the original and the recalibrated model were similar (AUC = 0.73, 95% confidence interval 0.67–0.78). However, significant miscalibration was found when the original QUANTEC model was tested in the current population, with a higher number of observed versus predicted events (HL-test p-value < 0.001). After recalibration, the HL-test showed a non-significant deviation (p-value 0.29) (Table 2a and Fig. 2), indicating no significant differences between observed and predicted incidences of RP. The (significant) HL-test shows that the original model was not consistent with the observed data, but that the final model is. Other performance measures can be found in the supplementary materials tables 2 and 3.

Ridge regression was used to prevent overfitting. The resulting final QUANTEC model is shown in Fig. 2c and given by:

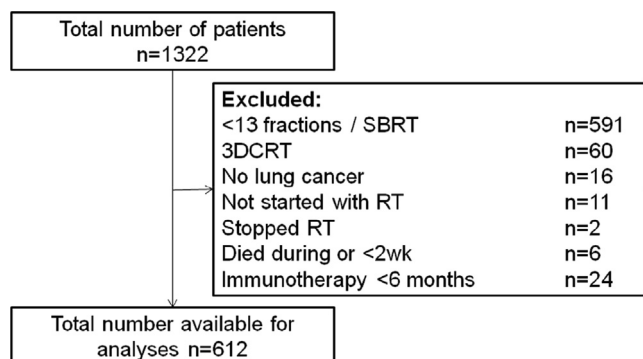$$NTCP_{finalQUANTEC} = (1 + \exp(4.575 - 0.224 * MLD))^{-1}$$

Table 2
A: External validation of the QUANTEC model; B: External validation of the APPELT model.

Variables	Original model	Recalibrated model	Final model	Variables	Original APPELT model	Revised model	Updated model	Final model
Model parameters				**Model parameters**				
Intercept	−3.870	−4.714	−4.575	Intercept	−4.760	−5.295	−8.303	−7.880
				MLD	0.138	0.247	0.267	0.250
MLD	0.126	0.234	0.224	Age (18–62 vs. > 63)	0.507	0.772		
				Age			0.052	0.049
				Smoking (previously Y/N)	−0.371	−0.107		
				Smoking (currently Y/N)	−0.478	−1.273		
				Smoking (current or quit < 3 months smoker vs never smoker or quit > 3 months)			−0.0939	−0.902
				Tumour location (lower lobe vs middle or upper lobe)	0.626	0.446		
				Pulmonary comorbidity	0.820	0.056		
				Chemotherapy (sequential Y/N)	0.470	−0.246		
Model performance				**Model performance**				
Discrimination:				Discrimination:				
AUC	0.725	0.725	0.725	AUC	0.692	0.774	0.787	0.787
Hosmer-Lemeshow test:				Hosmer-Lemeshow test:				
Chi squared	49.46	9.66	9.44	Chi squared	48.36	5.21	10.92	10.40
P-value	< 0.001	0.29	0.31	P-value	<0.001	0.735	0.206	0.238
Calibration				**Calibration**				
Intercept	2.472	0.000	0.064	Intercept	0.653	0.000	0.000	0.078
slope	1.857	1.000	1.045	Slope	1.006	1.000	1.000	1.060

Abbreviations 2a: Three models shown, the original QUANTEC model, the recalibrated model with an update of the intercept and MLD regression coefficient and the final model corrected for overfitting. MLD = Mean Lung Dose; CI = confidence interval; AUC = Area under the receiver operating characteristic curve. **2b:** Four models presented, the original APPELT model, the revised model with an update of all regression coefficients. the updated model after excluding the non-significant variables and after changing recoded variables and the final model corrected for overfitting. MLD = Mean Lung Dose; CI = confidence interval; AUC = Area under curve.

For the APPELT-model, the closed testing procedure recommended model revision, i.e., re-estimation of all regression coefficients (Table 2b). Following the model update, the intercept changed from −4.76 to −5.30, while the regression coefficient of the MLD increased from 0.14 to 0.25, indicating that MLD had more impact on the risk of RP than in the original APPELT-model. Additionally, the regression coefficient of age increased from 0.51 to 0.77, while the regression coefficients of the other variables decreased. After these adjustments, the AUC improved from 0.69 to 0.77.

Next, non-significant variables were eliminated stepwise, which resulted in exclusion of the following model parameters: pulmonary comorbidity, sequential chemotherapy (defined as sequential chemotherapy or no chemotherapy versus concurrent chemotherapy with or without induction chemotherapy) and tumour location. Additionally, the added value of regrouped or transformed variables was tested, resulting in a *new model* with MLD, age and smoking as predictors. This further improved the AUC to 0.79 (95% confidence interval 0.72–0.83) (Table 2b). Internal validation using bootstrapping, showed low optimism with a corrected slope of 0.95 (ideal 1.0). Detailed analyses can be found in the supplementary table 3 and 4.

Using Ridge regression, to correct for over fitting, the intercept shrunk from −8.303 to –7.880 (5.10%). This resulted in the *final updated APPELT model:*

$$NTCP_{FINAL} = (1 + \exp(7.880 - 0.250 * MLD + 0.902$$
$$* current\ Smoker\ or\ recently\ stopped$$
$$\langle 3months - 0.049 * Age))^{-1}$$

Calibration curves for this final updated APPELT-model are shown in Fig. 3. Calibration of the original model showed a higher number of observed versus predicted events (HL-test: p < 0.001).

The final model showed improved calibration (HL-test: p = 0.238) (Table 2b and Fig. 3). This resulted in the Final NTCP-model as shown in Fig. 4.

In comparison to the final QUANTEC-model, discrimination of the Final updated APPELT-model improved (AUC 0.79 vs. 0.73). The explained variance (Nagelkerke's R2) was 0.228 for this model, compared to 0.136 for the final QUANTEC-model, indicating that more variability was explained by the updated APPELT-model.

Discussion

In this prospective cohort study consisting of 612 lung cancer patients treated with recent photon radiation techniques (IMRT and VMAT), the most frequently used NTCP-models for RP were externally validated. This analysis revealed that the best performing model was the final updated model of Appelt et al., henceforth referred to as the New RP-model.

In line with previous studies, this analysis demonstrated that the risk of RP is not determined by MLD alone [1,17]. The individual sensitivity to a certain MLD is determined by co-factors, such as age and smoking which are currently included in the New-RP model (Fig. 4). Age and smoking, are well-known predictors for RP [18–19].

The QUANTEC- and the Appelt-model were selected for external validation. Other previously described models, such as of Huang and Reibnitz et al, were not selected because these included D_{max} parameters which are highly dependent of the prescribed dose and difficult to use in treatment planning optimization. The selected models were also externally validated in a retrospective cohort of 241 NSCLC patients by Thor et al [1–2,9,17,21]. In line with our findings, their analysis revealed that the original QUANTEC- and APPELT-models underestimated the risk of RP. The incidence of RP in our prospective cohort was 14.5% compared 35.9% in the APPELT cohort, which is probably related to a higher MLD (18.2 Gy compared to 11.5 Gy). However, compared with
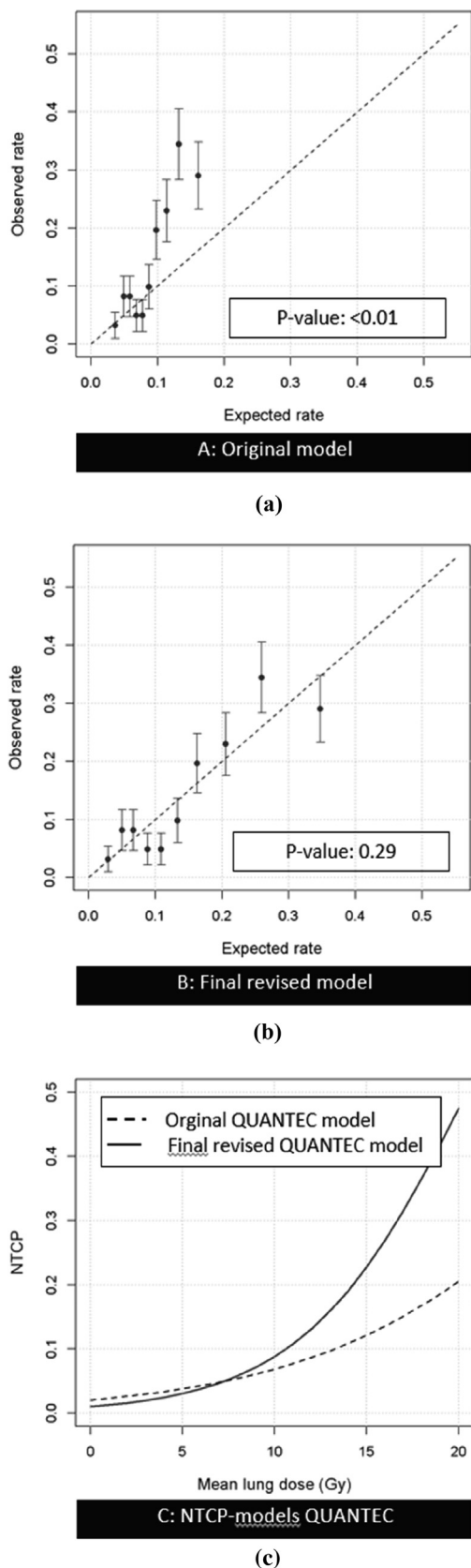
(a)

(b)

(c)

Fig. 2. Calibration plots and QUANTEC models A: calibration plot original model, B: calibration plot final revised model recalibrated and corrected for optimism, C: Original and final NTCP models QUANTEC.

the expected incidence based on the QUANTEC- or APPELT-model (8.8% and 8.5%, respectively), the incidence in our cohort was relatively high. This might be partly explained by our standardized and prospective scoring of RP but also by poor generalisation of the existing models to our cohort. In addition to the difference in RP rates, the incidence of pulmonary comorbidity in the Appelt-cohort was 19% compared to 36% in the current cohort. Despite this higher incidence, the regression coefficient was lower and not statistically significant and thus this variable was removed from the model. The portions of smoking patients and mean age were similar.

NTCP-models have become increasingly integrated in daily clinical practice to guide individualized treatment planning. These NTCP-models can also be used for model-based selection of lung cancer patients for more advanced techniques such as proton therapy [22]. Such clinical applications require reliable, high quality and generalizable NTCP-models [15].

The closed testing procedure is a well-accepted procedure in clinical epidemiology to determine whether an existing prediction model is applicable in an independent cohort, and indicates which components of the model should be adjusted or eliminated [10,15].

In the current analysis, the closed testing procedure indicated revision of both models. There are several reasons that may explain why these models required revision. Firstly, the QUANTEC model is a univariable model, including only the MLD as a predictor, which was based on the fitting of 10 different retrospective dataset [2]. It is very likely that the variation in these patient cohorts may have flattened out the dose–response curve. Secondly, these studies retrospectively assessed RP, in contrast to our standardized prospective scoring methodology, which may have resulted in misclassification of RP. Thirdly, the steepness of the recalibrated NTCP-model is also influenced by differences in treatment regimens. Marks et al. already argued that the QUANTEC model may not be valid for chemoradiotherapy and/or IMRT or VMAT [2]. Appelt included chemotherapy sequence as a variable in their model (sequential vs. concomitant chemotherapy). In the current analyses, 85.8% of the patients received any type of chemotherapy and included the non-chemotherapy group to have a representative cohort of current clinical practice. Chemotherapy increases the risk of RP, as it sensitizes not only the tumour but also normal tissues [23–24]. Some authors have suggested that gemcitabine, which was most frequently applied up to 2017 in our cohort (70.1% of the patient treated with concurrent chemotherapy), has a synergistic effect on radiation-induced toxicity [25–26]. Arietta et al. reported an exceptionally high rate of RP (31.5% $\geq$ grade 3) after induction gemcitabine / carboplatin followed by concurrent gemcitabine [26]. In another cohort, 30.2% grade 2 or higher toxicity was reported after induction gemcitabine follow by concurrent gemcitabine / paclitaxel [25]. In our institute, the RP rate ($\geq$grade 2) amongst patients who received concurrent gemcitabine, was 19.7%, which is in line with our previous report including an older cohort of patients (2002–2008) [27]. Moreover, additional analyses in our patient cohort showed that gemcitabine did not modulate the effect of the MLD on the risk of RP (see the supplementary material table 5). Furthermore, additional analyses were performed to investigate the possible effect of different chemotherapy schedules (induction plus concurrent, concurrent only, sequential and no chemotherapy). To this end, the original sequential chemotherapy variable (sequential chemotherapy or no chemotherapy versus concurrent chemotherapy with or without induction chemotherapy) was regrouped in 1) any chemotherapy yes / no and 2) and sequential chemotherapy yes / no (sequential vs concurrent chemotherapy with or without induction chemotherapy and no chemotherapy) and replaced in the analyses. None of the variables remained a predictor after exclusion of non-significant variables (see the supplementary material table 4). A
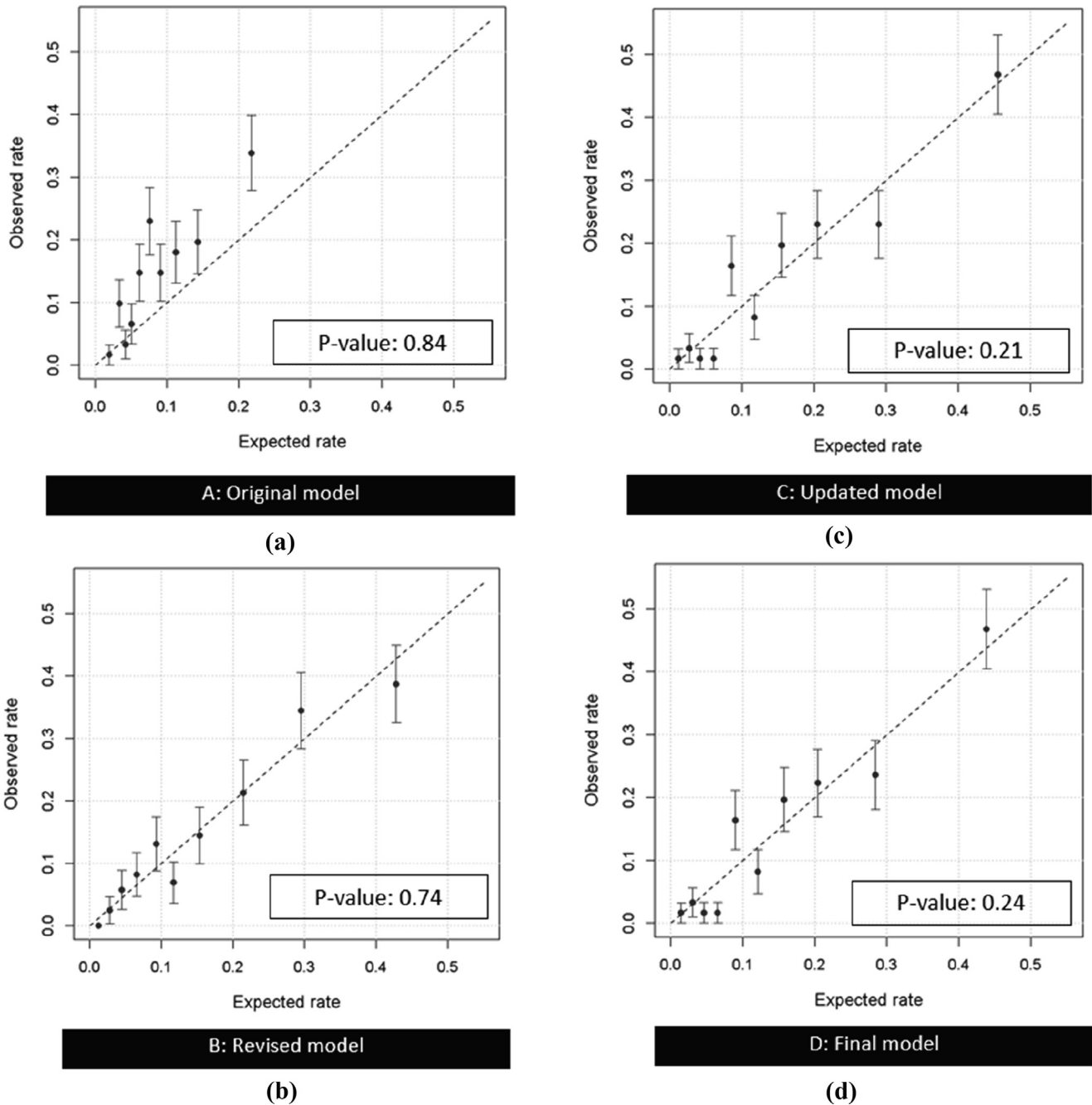
(a)

(c)

(b)

(d)

Fig. 3. Calibration plots of the original, revised, updated and final model. A: the original model; B: the revised model with an update of all the regression coefficients; C: the updated model the excluding the non-significant variables and the changed recoded variables; D: the final model corrected for overfitting.

more detailed analysis including the use different chemotherapy schedules and regimens was out of the scope of this analysis but might be of importance in future research as well as the influence of immunotherapy.

The use of advanced radiation techniques with consequently re-distribution of the dose, might also affect the risk of RP. In the current external validation, only IMRT and VMAT were used in contrast to 3D-CRT in the QUANTEC and APPELT analysis. However, in contrast to what we expected, the RP rate in the current cohort was higher than expected based on these (3D-CRT based) models. This might be explained by the larger volume of the lungs receiving 5 Gy or more [28]. Appelt et al. did not evaluate the effect of DVH parameters other than the MLD. However, the question arises whether the MLD is still the most suitable DVH parameter to predict the risk of RP. Several studies have shown the importance of

other lung DVH parameters, such as the lung V20 [19–20] or the lung V30 [29]. Huang et al. and Reibnitz et al. included D_{max} parameters which are, however, highly dependent on the prescribed dose and difficult to use in treatment planning optimization [17,21]. In contrast to photon radiotherapy, proton radiotherapy results in dose reductions over almost the entire dose range. A previously published external validation study using proton therapy data showed that the QUANTEC and APPELT models were also valid with minor modification for this modality [30]. Additional external validation is also required for different fractionation doses, such as used in stereotactic body radiotherapy (SBRT).

A limitation of our study is that we did not include patients treated with immunotherapy. Adjuvant immunotherapy has recently become the new standard of care in the treatment of
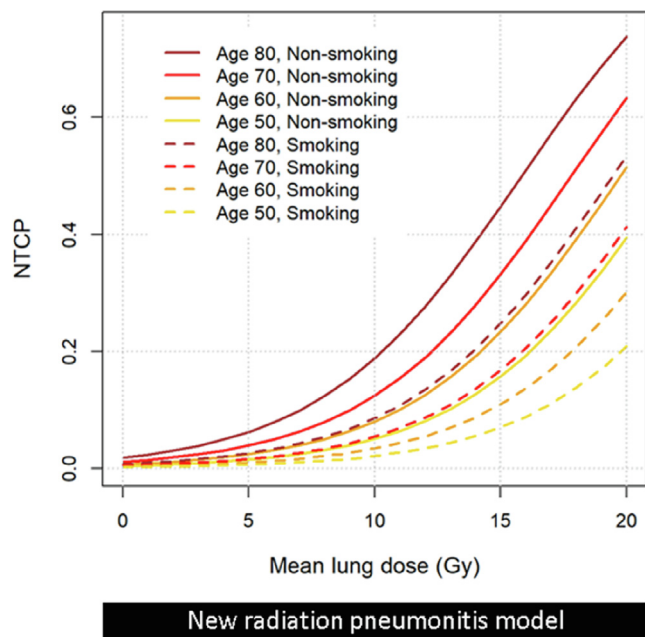
Fig. 4. Final NTCP-model for radiation pneumonitis.

locally advanced NSCLC [31–32]. Recent studies suggested that immunotherapy increases the risk of RP [33]. To keep models up to date it is important to evaluate the influence of new treatment techniques and chemotherapy and immunotherapy regimens will be evaluated in a future studies, but requires inclusion of more patients with longer follow up. As the New RP model contains only non-tumour specific parameters it could also be used for external validation in other patient cohorts.

Conclusion

Our external validation demonstrated that the APPELT model for RP, including clinical variables, performed better than the QUANTEC model. Nevertheless, model revision and update of the APPELT model were required, which resulted in the presented the New RP model for patients treated with modern photons-based radiation techniques like IMRT and VMAT. This model can be used to minimize the risk of RP, by guiding radiotherapy treatment planning or by selection of patients for new radiotherapy technologies.

Data sharing statement

Data from the analysis in this manuscript may be available to other researchers on request in accordance with institutional general policies. For more information, contact the corresponding author of this paper.

Disclosures

Declaration of Competing Interest

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

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2023.109735.

References

[1] Appelt AL, Vogelius IR, Farr KP, Khalil AA, Bentzen SM. Towards individualized dose constraints: Adjusting the QUANTEC radiation pneumonitis model for clinical risk factors. Acta Oncol (Madr) 2014;53:605–12.

[2] Marks LB et al. Radiation dose-volume effects in the lung. Int J Radiat Oncol Biol Phys 2010. https://doi.org/10.1016/j.ijrobp.2009.06.091.

[3] Graham MV et al. Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys 1999. https://doi.org/10.1016/S0360-3016(99)00183-2.

[4] Langendijk JA et al. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013. https://doi.org/10.1016/j.radonc.2013.05.007.

[5] Widder J et al. The quest for evidence for proton therapy: model-based approach and precision medicine. Int J Radiat Oncol Biol Phys 2016. https://doi.org/10.1016/j.ijrobp.2015.10.004.

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

[7] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63.

[8] Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: Validating a prognostic model. BMJ 2009. https://doi.org/10.1136/bmj.b605.

[9] Thor M et al. Toward personalized dose-prescription in locally advanced non-small cell lung cancer: Validation of published normal tissue complication probability models. Radiother Oncol 2019;138:45–51.

[10] Vergouwe Y et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 2017. https://doi.org/10.1002/sim.7179.

[11] Bentzen SM et al. Normal tissue effects: Reporting and analysis. Semin Radiat Oncol 2003. https://doi.org/10.1016/S1053-4296(03)00036-5.

[12] Niezink AGH et al. An instrument dedicated for modelling of pulmonary radiotherapy. Radiother Oncol 2015;115:3–8.

[13] Marks LB et al. Radiation dose-volume effects in the lung. Int J Radiat Oncol ● Biol ● Phys 2015;76:S20–7.

[14] van Buuren S, Oudshoorn CGM. MICE: multivariate imputation by chained equations. R Packag version 2007;1:2007.

[15] Van den Bosch L et al. Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy. Radiother Oncol 2020;148:151–6.

[16] Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley and Sons; 2004.

[17] von Reibnitz D et al. Predictive modeling of thoracic radiotherapy toxicity and the potential role of serum alpha-2-macroglobulin. Front Oncol 2020;10:1–9.

[18] Vogelius IR, Bentzen SM. A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis. Acta Oncol 2012. https://doi.org/10.3109/0284186X.2012.718093.

[19] Luna JM et al. Predicting radiation pneumonitis in locally advanced stage II–III non-small cell lung cancer using machine learning. Radiother Oncol 2019;133:106–12.

[20] Palma DA et al. Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: An international individual patient data meta-analysis. Int J Radiat Oncol Biol Phys 2013;85:444–50.

[21] Huang EX et al. Heart irradiation as a risk factor for radiation pneumonitis. Acta Oncol (Madr) 2011. https://doi.org/10.3109/0284186X.2010.521192.

[22] Langendijk JA et al. Clinical trial strategies to compare protons with photons. Semin Radiat Oncol 2018;28:79–87.

[23] Parashar B et al. Chemotherapy significantly increases the risk of radiation pneumonitis in radiation therapy of advanced lung cancer. Am J Clin Oncol Cancer Clin Trials 2011;34:160–4.

[24] Citrin DE, Mitchell JB. Altering the response to radiation: Sensitizers and protectors. Semin Oncol 2014. https://doi.org/10.1053/j.seminoncol.2014.09.013.

[25] Guilbault C et al. Long-term outcomes of induction carboplatin and gemcitabine followed by concurrent radiotherapy with low-dose paclitaxel and gemcitabine for stage III non–small-cell lung cancer. Clin Lung Cancer 2017;18:565–71.

[26] Arrieta O et al. High frequency of radiation pneumonitis in patients with locally advanced non-small cell lung cancer treated with concurrent

radiotherapy and gemcitabine after induction with gemcitabine and carboplatin. J Thorac Oncol 2009;4:845–52.

[27] Kerner GSMA et al. Concurrent gemcitabine and 3D radiotherapy in patients with stage III unresectable non-small cell lung cancer. Radiat Oncol 2014;9:190.

[28] Jo IY et al. Significance of low-dose radiation distribution in development of radiation pneumonitis after helical-tomotherapy-based hypofractionated radiotherapy for pulmonary metastases. J Radiat Res 2014;55:105–12.

[29] Sheng L, Cui X, Cheng L, Chen Y, Du X. Risk factors of grade ≥ 2 radiation pneumonitis after gemcitabine induction chemotherapy for patients with non-small cell lung cancer. Radiat Oncol 2019;14:1–9.

[30] Jain V et al. Updating photon-based normal tissue complication probability models for pneumonitis in patients with lung cancer treated with proton beam therapy. Pract Radiat Oncol 2020. https://doi.org/10.1016/j.prro.2020.04.005.

[31] Antonia SJ et al. Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer. N Engl J Med 2017;377:1919–29.

[32] Faivre-Finn C et al. Four-year survival with durvalumab after chemoradiotherapy in stage III NSCLC—an update from the PACIFIC trial. J Thorac Oncol 2021;16:860–7.

[33] Jung HA et al. Real world data of durvalumab consolidation after chemoradiotherapy in stage III non-small-cell lung cancer. Lung Cancer 2020;146:23–9.