

A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: A step toward individualized care and shared decision making

Citation for published version (APA):

Oberije, C., Nalbantov, G., Dekker, A., Boersma, L., Borger, J., Reymen, B., van Baardwijk, A., Wanders, R., De Ruyscher, D., Steyerberg, E., Dingemans, A.-M., & Lambin, P. (2014). A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: A step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112(1), 37-43. <https://doi.org/10.1016/j.radonc.2014.04.012>

Document status and date:

Published: 01/07/2014

DOI:

[10.1016/j.radonc.2014.04.012](https://doi.org/10.1016/j.radonc.2014.04.012)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 20 Apr. 2024



Model based prediction

A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: A step toward individualized care and shared decision making [☆]



Cary Oberije ^{a,*}, Georgi Nalbantov ^a, Andre Dekker ^a, Liesbeth Boersma ^a, Jacques Borger ^a, Bart Reymen ^a, Angela van Baardwijk ^a, Rinus Wanders ^a, Dirk De Ruyscher ^b, Ewout Steyerberg ^c, Anne-Marie Dingemans ^d, Philippe Lambin ^a

^a Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, The Netherlands; ^b Department of Radiation Oncology, University Hospital Leuven/KU Leuven, Belgium; ^c Department of Public Health, Erasmus Medical Center, Rotterdam; and ^d Department of Pulmonology, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, The Netherlands

ARTICLE INFO

Article history:

Received 6 February 2014
Received in revised form 14 April 2014
Accepted 18 April 2014
Available online 17 May 2014

Keywords:

Prediction models
Lung cancer

ABSTRACT

Background: Decision Support Systems, based on statistical prediction models, have the potential to change the way medicine is being practiced, but their application is currently hampered by the astonishing lack of impact studies. Showing the theoretical benefit of using these models could stimulate conductance of such studies. In addition, it would pave the way for developing more advanced models, based on genomics, proteomics and imaging information, to further improve the performance of the models.

Purpose: In this prospective single-center study, previously developed and validated statistical models were used to predict the two-year survival (2yrS), dyspnea (DPN), and dysphagia (DPH) outcomes for lung cancer patients treated with chemo radiation. These predictions were compared to probabilities provided by doctors and guideline-based recommendations currently used. We hypothesized that model predictions would significantly outperform predictions from doctors.

Materials and methods: Experienced radiation oncologists (ROs) predicted all outcomes at two timepoints: (1) after the first consultation of the patient, and (2) after the radiation treatment plan was made. Differences in the performances of doctors and models were assessed using Area Under the Curve (AUC) analysis.

Results: A total number of 155 patients were included. At timepoint #1 the differences in AUCs between the ROs and the models were 0.15, 0.17, and 0.20 (for 2yrS, DPN, and DPH, respectively), with *p*-values of 0.02, 0.07, and 0.03. Comparable differences at timepoint #2 were not statistically significant due to the limited number of patients. Comparison to guideline-based recommendations also favored the models.

Conclusion: The models substantially outperformed ROs' predictions and guideline-based recommendations currently used in clinical practice. Identification of risk groups on the basis of the models facilitates individualized treatment, and should be further investigated in clinical impact studies.

© 2014 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 112 (2014) 37–43

Background

It is widely recognized that clinical decision making is not straightforward for most oncological treatments. Diagnostic procedures provide an increasing amount of information, and the

number of treatment options for individual patients are also growing. In addition, patient preferences can differ considerably and should also be taken into account.

Decision Support Systems (DSSs) can provide clinicians and patients with patient-specific information about which patients are most likely to benefit from a given treatment and which ones are most likely to experience the harmful side effects of a treatment. Although their development and use are still in their infancy, there is an increasing interest for using DSSs in medicine and, specifically, in oncology [1]. DSSs have the potential to greatly impact patient management and clinical practice. Individualized prognos-

An abstract related to this study has been presented at the annual meeting of the European Society of Therapeutic Radiology and Oncology (ESTRO) (April, 2013).

* Corresponding author. Address: MAASTRO Clinic, Dr. Tanslaan 12, 6229 ET Maastricht, The Netherlands.

E-mail address: cary.oberije@maastro.nl (C. Oberije).

<http://dx.doi.org/10.1016/j.radonc.2014.04.012>

0167-8140/© 2014 Elsevier Ireland Ltd. All rights reserved.

tic information will not only stimulate shared decision making, but also facilitate adaptation of treatment, allow for more rational choices between treatment options, and stimulate innovations of clinical trial design. Figs. 1 and S1 show clinical applications in which active patient participation, treatment adaptation, and treatment choice are integrated in a decision tree and guided by predictive models. However, considering the abundant number of published predictive models and the astonishing absence of impact studies, which are required to assess changes in patient management, a study that bridges this gap is needed [2]. Therefore, we decided to investigate whether patient-specific prognostic information obtained from predictive models has added value as compared to information provided by the treating doctor or guideline-based recommendations.

We previously developed and validated models to predict three important outcomes for lung cancer patients: overall survival, dyspnea and dysphagia [3–5]. In general, predictions based on statistical rules or models are at least as reliable as, and typically more reliable than, the predictions of human experts if based on the

same information [6]. Indeed, studies focusing on survival prediction of terminally ill cancer patients have shown that predictions of doctors tend to be too optimistic, unreproducible, and inaccurate [7–12]. For lung cancer patients, this information is not yet available. Also, the most widely used system to stratify lung cancer patients into risk categories, the TNM staging system, has limitations when used for lung cancer patients treated with chemoradiation [13]. Finally, studies investigating doctors' predictions of severe treatment-induced side-effects in lung cancer patients are currently lacking.

The purpose of this prospective study was, thus, to compare predictions for two-year death rate (2yrD), severe acute treatment-induced dysphagia and severe acute treatment-induced dyspnea of lung cancer patients, based on statistical models, to (A) predictions made by the radiation oncologists (ROs) after they had seen the patient, and (B) guideline-based recommendations. The models were previously developed and externally validated [3–5,14]. We hypothesized that these models would significantly outperform the prediction made by ROs.

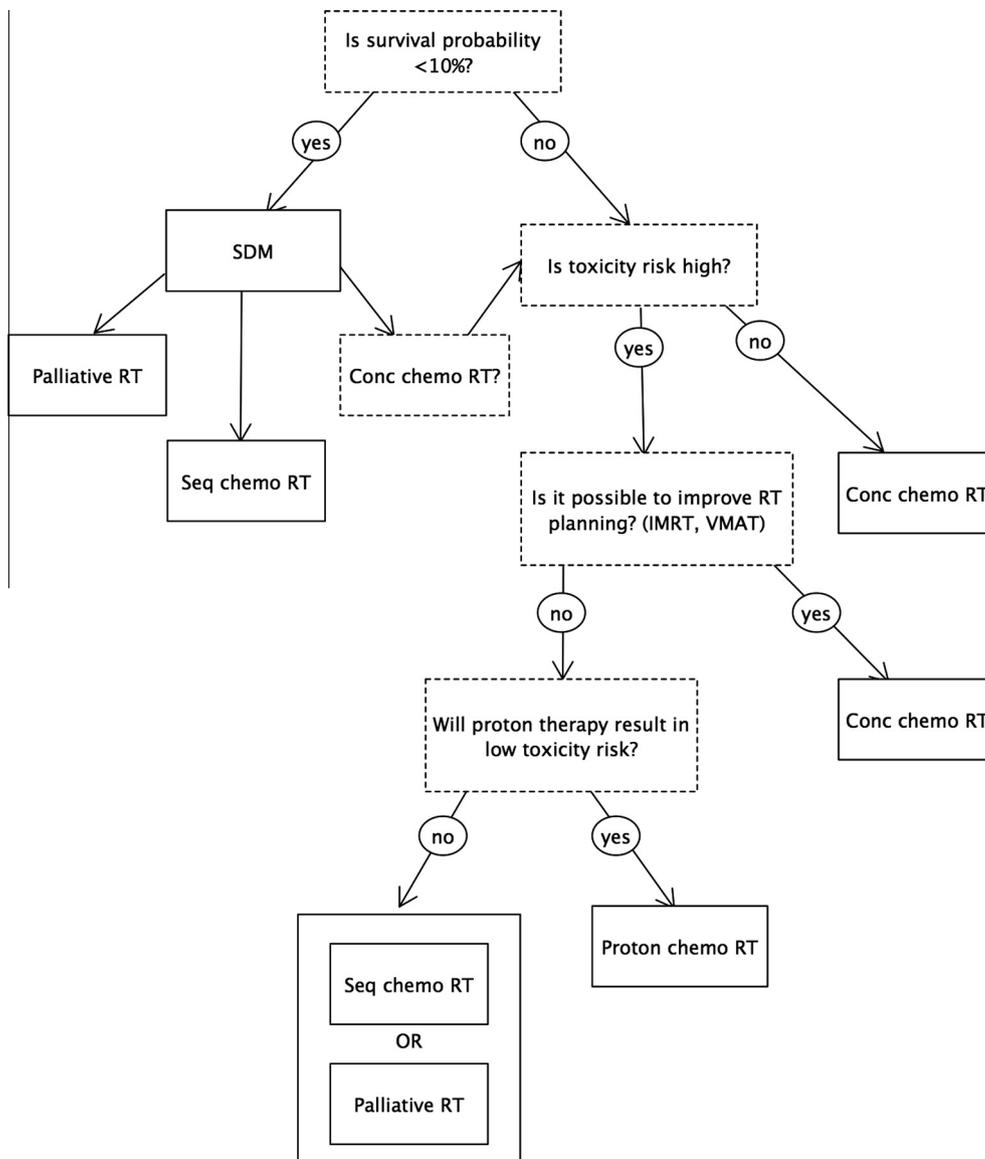


Fig. 1. Clinical application of a Decision Support System for stage III NSCLC patients. Shared decision making (SDM), adaptation of treatment (improved radiotherapy planning) and choice of treatment options (either concomitant or sequential chemo radiotherapy and palliative radiotherapy) are integrated in a decision tree based on previously developed and validated prognostic models.

Methods

Details of the study

Based on the performance of the previously published and validated predictive models, and the results of a pilot study in which the ROs were provided with only written information, we hypothesized that the models would outperform the ROs' predictions by at least 0.12 in Area Under the Curve (AUC) analyses. The maximum value of the AUC is 1.0; indicating a perfect prediction model. A value of 0.5 indicates that patients are correctly classified in 50% of the cases; i.e., the same as random chance. The primary outcome of the study was the prediction of two-year survival. To achieve an 80% power level, using a two-sided z-test at a significance level of 0.05, 129 predictions were required. It was decided to use only data from NSCLC patients for the survival outcome, as the NSCLC and the SCLC population are not comparable with respect to this outcome. However, for prediction of acute severe dysphagia and acute severe dyspnea also SCLC patients were included. Between July 2007 and May 2009, 129 patients with medically- or technically-inoperable stage I–IIIB non-small cell lung cancer (NSCLC), and 26 patients with stage I–IIIB small cell lung cancer (SCLC), all treated with curative intent by chemo radiotherapy, were entered into this prospective single-center study.

Treatment schedules

Patients were treated in accordance to different *a priori* defined radiation and chemotherapy protocols according to stage and histology (NSCLC vs. SCLC).

All patients received radiation treatment at the MAASTRO clinic; the chemotherapy treatment was administered in local hospitals. No elective nodal irradiation was performed [15,16]. Three different radiotherapy treatment regimes were administered. Details of the regimes have been published previously [17–19].

Description of the prediction models

The model for overall survival of NSCLC, developed on a dataset of 322 patients (AUC = 0.75) and externally validated (AUC = 0.76), consisted of five variables: sex, WHO performance status (WHO-PS), forced expiratory volume in one second (FEV₁), gross tumor volume (GTV), and number of positive lymph node stations (PLNS) [3].

Data from 407 patients were used to build a model for radiation-induced dyspnea (AUC = 0.62) [5]. Five variables were included in the model: age, WHO-PS, FEV₁, nicotine use, and mean lung dose (MLD). External validation was performed on data from 175 patients (AUC = 0.61) [14].

The model for prediction of treatment-induced dysphagia was developed on data from 469 patients (cross-validated C-statistic = 0.77) and included seven variables: sex, age, WHO-PS, overall treatment time (OTT), mean esophageal dose (MED), maximum esophageal dose (MAXED), and chemotherapy. External validation of two datasets yielded AUCs of 0.94 and 0.77 [4].

More detailed information can be found in the [Supporting information \(Text S2\)](#) or at our website www.predictcancer.org.

Prediction of the radiation oncologists

Radiation oncologists (ROs), all specialized in lung cancer with clinical experience ranging from 3 to >20 years, were asked to indicate the probability (expressed as a percentage) that a patient would survive for more than 2 years, suffer from severe treatment-induced dysphagia, or suffer from severe treatment-induced

dyspnea. The ROs based their predictions on all available information, including: medical history, radiology and pathology reports, PET–CT images, lung function test results, and physical examination. They filled in a prespecified form. Severe toxicity was defined as toxicity \geq grade III, according to the CTCv3. Assessments were made at two timepoints: (1) after a patient's first visit to the radiotherapy department, and (2) after the radiotherapy treatment plan had been made and thus the gross tumor volume (GTV) and the dose volume histogram (DVH) were available.

Guideline-based recommendations

The performances of models and doctors were also compared to currently used EORTC recommendations [20]. For survival, the TNM staging system was used to create risk groups. The risk of developing severe radiation-induced dyspnea is limited by constraining the mean lung dose [21]. According to the EORTC this dosimetric parameter should not exceed 20 Gy, which will restrict the incidence of severe radiation-induced dyspnea to 10–15%. In case of radiation-induced dysphagia the EORTC recommends that the maximal esophageal dose should be lower than 79 Gy [20]. Depending on the treatment, by constraining this dose, 5% (no chemotherapy), 20% (sequential chemotherapy), or 40% (concomitant chemotherapy) of patients are expected to develop severe dysphagia.

Outcome assessment

The primary outcome of the study was two-year survival. Survival was evaluated in October 2012 and defined as the duration between the start of radiation therapy and the date of death. Toxicity during and after treatment was scored by ROs using the dyspnea and dysphagia scales of the NCI Common Toxicity Criteria version 3.0 (CTCAEv3). For the analyses, the maximum dysphagia score at any timepoint during or after the end of radiotherapy, with a maximum of 4 weeks, was used. For dyspnea analyses, the maximum dyspnea score after radiotherapy, but within 6 months after the end of radiotherapy, was used.

Statistical analysis

The model coefficients of the previously developed, validated, and published models were used to calculate the probability for an individual patient for each outcome [3–5,14]. For all variables, single imputation based on predictive mean matching was used if a value was missing (additional information on multiple imputation analysis is provided in the Supporting information S3). This method ensures that the imputed values are plausible. The performances of models and ROs were compared with respect to calibration and discrimination. Calibration refers to the agreement between the observed frequencies and the predicted probabilities. Discriminative ability (i.e., whether the relative ranking of individual predictions was in the correct order) was determined by Area Under the Curve (AUC) analysis of the Receiver Operator Curve (ROC). In addition, for the ROs as well as for the models, the median value of the predicted probability was used to split the predicted probability variable into two groups. For this specific cut point the accuracy, sensitivity, specificity, and positive (PPV) and negative predictive values (NPV) were calculated. For the survival analysis only data of NSCLC patients were used. Kaplan–Meier curves were made based on the clinical TNM stage or risk groups. Low-, medium-, and high-risk groups were created by splitting the patient group into three groups, based on the 25th and 75th percentile points of the predicted probabilities [22].

The analysis was performed with SPSS, version 19.11 (SPSS Inc., Chicago, IL) and R, version 2.14.0 (R foundation for statistical computing, Vienna, Austria).

Ethics

This prospective study was conducted according to national laws and guidelines, approved by the appropriate Institutional Review Board. Clinical data were collected according to prospective trials, that were registered with ClinicalTrials.gov (NCT00572325, NCT00181545 and NCT00572923).

Results

A total of 155 patients were entered in the study (Table 1). For the survival outcome only data of the NSCLC patients were used. Analyzing the data of the 129 NSCLC patients, it was found that two patients had lesions in multiple lung lobes, one patient was treated for recurrent disease and five patients had oligometastases. These eight patients were included for the analysis of the dyspnea and dysphagia outcome, but they were excluded from the survival comparison because they did not meet the inclusion criteria. To assess the robustness of the results, an additional intention-to-treat analysis was performed by analyzing all 129 patients. If one of the outcomes was not predicted, the patient was only excluded for that specific outcome analysis. Consequently, the analyses for dysphagia, dyspnea, and survival were based on the data from 146, 139, and 121 patients, respectively. The number of predictions

per RO is shown in Table S4. For timepoint #2 predictions for only 45 patients were available, resulting in 42, 40, and 36 patients included for dysphagia, dyspnea and survival outcome, respectively.

The models outperformed the doctors as well as the guideline-based recommendations, resulting in higher performance measures (Fig. 2). The AUCs of the ROC curves were 0.52, 0.59 and 0.56 for the doctors, while the models yielded AUCs of 0.71, 0.76 and 0.72 (p -values 0.02, 0.07, and 0.03, respectively; Fig. S5; Table 2). The KM curves based on clinical TNM stage or ROs' predictions could not identify survival risk groups (logrank test: $p = 0.33$ and $p = 0.29$, Fig. 3a and b), in contrast with the KM curves based on model predictions (logrank test: $p = 0.001$, Fig. 3c). The difference between the predicted probability for patients with or without the event of interest was larger for the models than for the ROs (Fig. S6). The models were thus more capable to discriminate

Table 1
Patient characteristics.

		NSCLC (N = 129)	SCLC (N = 26)
Sex	Male	75 (58.1%)	11 (42.3%)
	Female	54 (41.9%)	15 (57.7%)
Age (years)	Mean (SD)	64.7 (10.5)	65.5 (8.8)
	WHO-PS		
WHO-PS	0	59 (46.8%)	9 (34.6%)
	1	60 (47.6%)	11 (42.3%)
	≥2	7 (5.6%)	6 (23.1%)
FEV ₁ (%)	Mean (SD)	78 (22)	71 (20)
	Nicotine use		
Nicotine use	Current smoker	47 (38.8%)	8 (33.3%)
	No/ex smoker	74 (61.2%)	16 (66.7%)
cTNM	I	17 (13.4%)	–
	II	3 (2.4%)	4 (15.4%)
	IIIA	48 (37.8%)	7 (26.9%)
	IIIB	50 (39.4%)	15 (57.7%)
	IV	5 (4.0%)	–
PLNS	0	27 (22.9%)	NA
	1	19 (16.1%)	
	2	31 (26.3%)	
	3	14 (11.9%)	
	≥4	27 (22.9%)	
GTV (cc)	Median (range)	72 (0–763)	62 (8–398)
	Chemotherapy		
Chemotherapy	No	24 (18.6%)	1 (3.8%)
	Sequential	31 (24.1%)	2 (7.7%)
	Concomittant	74 (57.4%)	23 (88.5%)
	Mean (SD)	61.6 (7.7)	46.6
TTD (Gy)	Mean (SD)	28 (7)	22 (4)
OTT (days)	Mean (SD)	24.4 (10.8)	20.8 (9.3)
MED (Gy)	Mean (SD)	57.7 (14.1)	44.3 (7.6)
MAXED (Gy)	Mean (SD)	16.0 (3.4)	14.1 (3.3)
MLD (Gy)	Mean (SD)	64 (52.9%)	NA
Dead at 2 years		14 (11.6%)	3 (12.0%)
Dysphagia ≥grade 3		10 (8.8%)	1 (4.8%)
Dyspnea ≥grade 3			

Abbreviations: SCLC = small cell lung cancer; NSCLC = non-small cell lung cancer; WHO-PS = World Health Organization performance status; FEV₁ = forced expiratory volume (1 s); PLNS = number of positive lymph node stations; GTV = gross tumor volume (primary + nodes); TTD = total radiation dose; OTT = overall treatment time; MED = mean esophageal dose; MAXED = maximum esophageal dose; NA = not assessed.

* Standard dose 45.0 Gy, two patients were not treated per protocol.

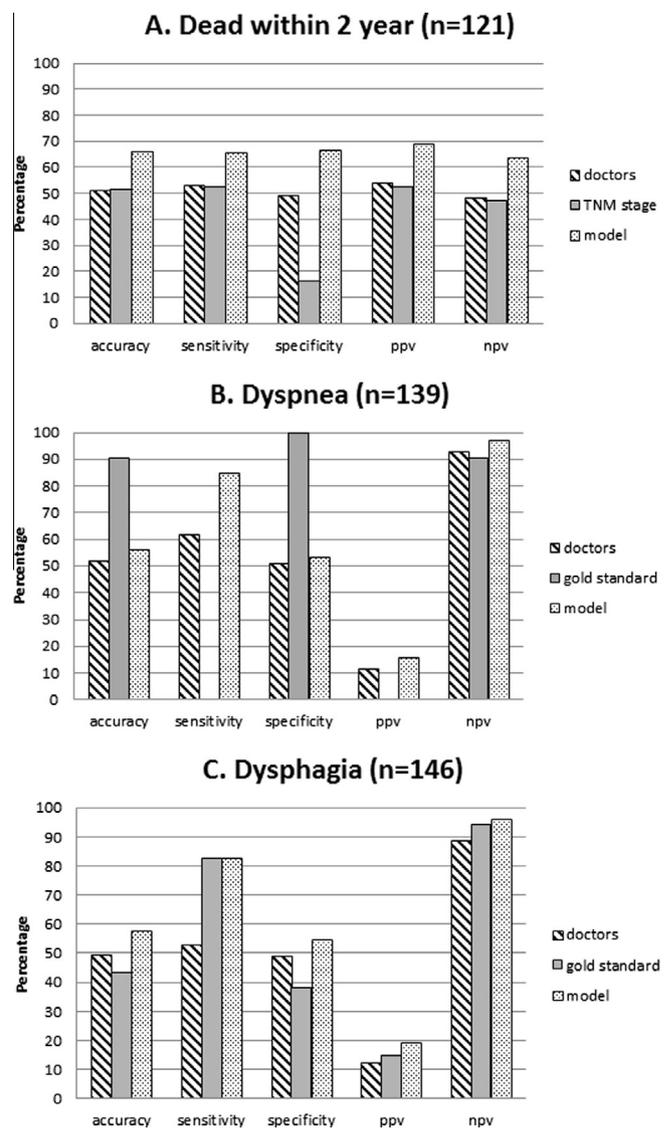


Fig. 2. Performance of the models compared to the doctors' or guideline-based recommendations for (A) death within two years, (B) severe treatment-induced dyspnea, and (C) severe treatment-induced dysphagia. The TNM staging system is often applied as a prognostic tool for survival; for dyspnea and dysphagia, the mean lung radiation dose and maximal esophageal radiation dose were used, respectively. The median value was used to dichotomize the predicted probabilities. For the dyspnea outcome splitting the group was not possible using the gold standard, as the predicted risk was the same for all patients.

Table 2
Comparison of doctors' versus models' predictions.

	Outcome	Doctors'		Model		N	p [*]
		AUC	95% CI	AUC	95% CI		
Timepoint 1	Dead within 2 years	0.56	0.46–0.67	0.71	0.61–0.80	121	0.02
	Dyspnea	0.59	0.44–0.74	0.76	0.62–0.90	139	0.07
	Dysphagia	0.52	0.39–0.66	0.72	0.59–0.85	146	0.03
Timepoint 2	Dead within 2 years	0.56	0.36–0.75	0.78	0.62–0.93	35	0.05
	Dyspnea	0.61	0.35–0.88	0.74	0.44–1	39	0.51
	Dysphagia	0.64	0.34–0.83	0.87	0.61–1	41	0.32

* p-Value assessed with DeLong's test for two correlated ROC curves.

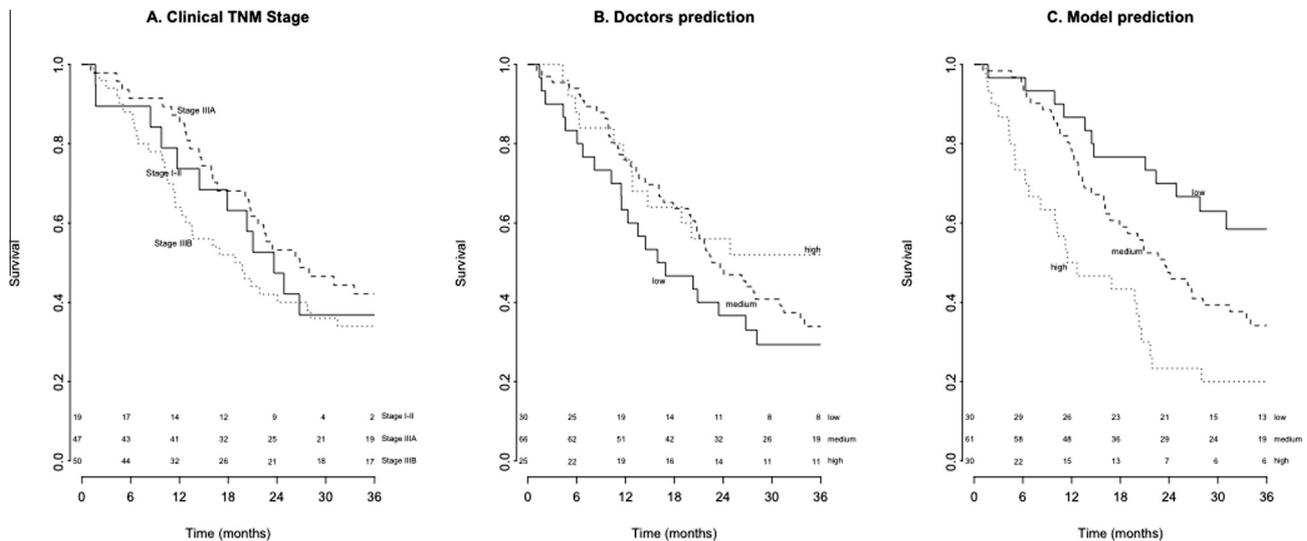


Fig. 3. Kaplan–Meier curves for overall survival (months) of 121 NSCLC patients based on (A) clinical TNM stage, (B) predictions of radiation oncologists, and (C) model predictions. *p*-Values of the logrank tests were 0.33, 0.29, and 0.001, respectively.

between patients with and without the event of interest. For most cut-off points, the models' NPV and PPV were higher (Fig. S7). Thus, applying a threshold value to identify either high-risk or low-risk patients based on model predictions will result in more accurate stratification if the model predictions are used instead of ROs prediction. The intention-to-treat analysis for two-year OS ($N = 129$ NSCLC patients) resulted in an AUC of 0.53 (CI 95%: 0.43–0.63) for the ROs, while the model achieved an AUC of 0.71 (CI 95%: 0.62–0.80, $p = 0.03$). The ROs overestimated the risk of all outcomes, whereas the models overestimated death and dysphagia risk, but underestimated the risk of dyspnea (Table S8). At timepoint #2 the differences between the models and the doctors were comparable to those at timepoint #1 (Table 2) but, not statistically significant due to the limited number of patients.

Discussion

To our knowledge, this is the first prospective study comparing the predictions of models versus doctors for survival and toxicity outcome in lung cancer patients. The models for two-year survival, severe dysphagia, and severe dyspnea significantly outperformed the predictions of the doctors as well as guideline-based recommendations. Moreover, the difference between the ROs and the models did not decrease after they had seen the radiotherapy plan. In our opinion, these results pave the way for the conductance of large-scale impact studies in which changes in patient management can be quantified.

Several studies have concluded that the predictive capability of doctors is limited for terminally ill cancer patients [7–10,23–26], as well as for prostate cancer patients treated with curative intent

[11,12]. This is consistent with our findings, which showed that the same may be concluded for NSCLC patients treated with curative chemo radiotherapy.

Moreover, we prospectively compared the doctors' predictions to previously published statistical models, thereby showing that the doctors' capabilities are not only limited, but also considerably worse than those of the models.

It might seem surprising that experienced ROs, who had access to all available data, including imaging scans and tumor volume, were not able to estimate the survival probability accurately. Indeed, tumor volume (GTV) has consistently been reported as a prognostic factor for survival of NSCLC patients [17,27–30] but its interpretation for individual patients is not straightforward. It has been shown that the impact of GTV on locoregional control can be influenced by the radiation fractionation schedule [29]. Moreover, there is no linear relationship between GTV and overall survival, which is reflected in the very limited prognostic value of GTV for tumors with a diameter larger than 5 cm [3,27,28].

In a review it was concluded that clinicians might benefit from assisted life expectancy predictions using life tables and statistical tools, as the accuracy of clinician-derived predictions for life expectancy in prostate cancer was relatively modest [11]. Moreover, an experiment in which urologists were allowed to adapt the nomogram prediction of five-year recurrence-free survival rates for prostate cancer patients showed that their modifications led to a significant decrease in prediction accuracy (c-index decreased from 0.67 to 0.55, $p < 0.05$) [12]. It has also been reported that prediction of survival is more difficult as survival time increases [9,26]. This so-called "horizon" effect might have influenced the predictions of the ROs, as we asked them to predict two-year

survival outcome. It cannot be excluded that they might be better at predicting short-term survival. However, for patients with NSCLC, treated with curative intent, two-year survival is a very relevant outcome.

The AUC was used for statistical comparison, but for clinical practice the negative (NPV) and positive predictive values (PPV) are more useful. We depicted these performance measures for all thresholds (Fig. S7) as well as for the median value (Fig. 2). The median value was an arbitrary choice, mainly meant to illustrate differences in performance in an understandable way. In clinical practice more relevant cut points may be chosen to minimize either false positive or false negative classification.

The model predictions were compared to EORTC based recommendations [20], which are used in many hospitals, but it cannot be excluded that guidelines based on other parameters or combinations of parameters might predict outcomes more accurately.

We applied prediction models that were previously developed at our institute and validated on external datasets. Recently, new prognostic factors, which could improve our models for dyspnea and dysphagia, have been published [31,32]. However, as data were prospectively collected, it is not possible to test these new variables on the patient cohort included in the current study.

Since limiting the mean lung radiation dose results in a low risk for developing dyspnea, all patients were categorized as low risk. Subsequently, the sensitivity of this dosimetric parameter was 0%, while the specificity was 100%. Thus, selection of high-risk patients based on this dosimetric parameter was not possible.

The number of predictions per RO varied considerably with 65% being predicted by two ROs (Table S1), who have each more than 20 years experience in treating lung cancer. Even if this is not a representative sample of the RO population in general, it can therefore be assumed that their performance is at least comparable to other ROs.

At timepoint #2 a limited number of predictions were available for analysis. This was mainly due to the working process in our clinic, where few patients are seen by the RO in the interval between planning and start of radiotherapy. Therefore, analyses of predictions at timepoint #2 resulted in statistically non-significant *p*-values. However, the absolute difference between the predictions of the doctors and models, in terms of the AUC, did not change.

In the medical domain, a number of biases can affect the ways in which doctors gather and use evidence in making diagnoses, prognoses, and treatment decisions, leading to over- or under-treatment of patients [33]. Our models, although not perfect, can offer valuable assistance in clinical decision making. By choosing cut-off points based on the model predictions we are able to define clinically-relevant low- or high-risk groups (Fig. S7). The PPV and NPV are comparable to widely accepted clinical tests. For example, one of the most often used screening tests, PSA level for prostate cancer, has a PPV of 30% and an NPV of 85% [34–36]. In addition, a consortium of leading radiation oncology experts has developed a model-based approach to select patients for proton therapy [37]. This methodology has been accepted by the Dutch health authorities despite the limited predictive performance of normal tissue complication probability (NTCP) models [38]. Taking into account the finite resources and the growing costs of health care, selecting subgroups of patients for specific treatment options, instead of delivering expensive treatments to the entire patient population, will be inevitable.

Our strategy for developing a Decision Support System consists of two main pillars. First, we are aiming to improve models by adding information based on image analysis, genomics and proteomics. Second, clinical impact studies, preferably randomized, will be initiated. In these studies, model based predictions will be used in the experimental arm to identify a subgroup of patients, for

whom consequently the treatment might be adapted. For instance, patients at very low risk for developing severe dyspnea could receive a higher radiation dose; patients at very high risk for death could go through a shared decision process to decide which treatment they prefer.

Genomics, proteomics and radiomics give rise to unprecedented increases in the amount of information available for an individual patient. The number of therapeutic options is growing as well, thereby complicating the tasks of doctors even further [39–41]. In a recent review we stated that Decision Support Systems (DSSs), integrating multiple models for a number of relevant outcomes, will eventually be considered as medical devices and used in daily clinical practice, analogous to radiotherapy treatment-planning software [1]. Methods to combine multiple models, that predict the same outcome, so-called “ensemble learning”, are currently being investigated and are not yet often applied in the medical domain. However, they offer new ways to integrate more and diverse information in models, which will eventually improve clinical predictions.

Development of DSSs will take time, but the first steps have already been taken. In our study, models based on general patient, tumor, and treatment characteristics outperformed the doctors’ predictions as well as guideline-based recommendations. In the near future, improvement of the models can be achieved by incorporating “-omics” information [39–41], but the human capacity to select and combine information will not change. Bearing in mind that current models already show superior performance, it seems that we are facing a new era, where it may eventually be regarded as unethical to base medical decisions solely on a doctor’s opinion.

Conflict of interest statement

Prof. De Ruyscher has declared the following relevant financial activities outside the submitted work: board membership of Astellas and Siemens, (pending) grants of BMS, Boehringer and Siemens. No actual or potential conflicts of interest exist for the other authors of this manuscript.

Funding

Funding for this research was provided by the Center for Translational Molecular Medicine (CTMM), the AirForce project; EFRO (European Fund for Regional Development), the EuroCAT project, and the Dutch Cancer Society (KWF UM 2011-5020 and KWF UM 2009-4454).

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.2014.04.012>.

References

- [1] Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol* 2012;10:27–40.
- [2] Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.
- [3] Dehing-Oberije C, Yu S, De Ruyscher D, Meersschout S, Van Beek K, Lievens Y, et al. Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2009;74:355–62.
- [4] Dehing-Oberije C, De Ruyscher D, Petit S, Van Meerbeeck J, Vandecasteele K, De Neve W, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010;97:455–61.

- [5] Dehing-Oberije C, De Ruyscher D, van Baardwijk A, Yu S, Rao B, Lambin P. The importance of patient characteristics for the prediction of radiation-induced lung toxicity. *Radiother Oncol* 2009;91:421–6.
- [6] Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychol Public Policy Law* 1996;2:293.
- [7] Clément-Duchêne C, Carnin C, Guillemin F, Martinet Y. How accurate are physicians in the prediction of patient survival in advanced lung cancer? *Oncologist* 2010;15:782–9.
- [8] Glare P, Virik K, Jones M, Hudson M, Eychmuller S, Simes J, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ* 2003;327:195.
- [9] Mackillop WJ, Quirt CF. Measuring the accuracy of prognostic judgments in oncology. *J Clin Epidemiol* 1997;50:21–9.
- [10] Chow E, Davis L, Panzarella T, Hayter C, Szumacher E, Loblaw A, et al. Accuracy of survival prediction by palliative radiation oncologists. *Radiat Oncol Biol* 2005;61:870–3.
- [11] Jeldres C. Life expectancy estimation in prostate cancer patients. *Can Urol Assoc J* 2012;6:374–5.
- [12] Ross PL, Gerigk C, Gonen M, Yossepowitch O, Cagiannos I, Sogani PC, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002;20:82–8.
- [13] Ball D, Smith J, Wirth A, Mac Manus M. Failure of T stage to predict survival in patients with non-small-cell lung cancer treated by radiotherapy with or without concomitant chemotherapy. *Radiat Oncol Biol* 2002;54:1007–13.
- [14] Oberije C, Liao Z, De Ruyscher D, Tucker S, Lambin P. Development and external validation of a model for prediction of radiation-induced dyspnea: an approach combining clinical data with information from literature. *Int J Radiat Oncol Biol Phys* 2010;78:5528EP.
- [15] De Ruyscher D, Bremer R-H, Koppe F, Wanders S, van Haren E, Hochstenbag M, et al. Omission of elective node irradiation on basis of CT-scans in patients with limited disease small cell lung cancer: a phase II trial. *Radiother Oncol* 2006;80:307–12.
- [16] Senan S, De Ruyscher D, Giraud P, Mirimanoff R, Budach V. Radiotherapy Group of European Organization for Research and Treatment of Cancer. Literature-based recommendations for treatment planning and execution in high-dose radiotherapy for lung cancer. *Radiother Oncol* 2004;71:139–46.
- [17] van Baardwijk A, Wanders S, Boersma L, Borger J, Öllers M, Dingemans A-MC, et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. *J Clin Oncol* 2010;28:1380–6.
- [18] van Baardwijk AA, Reymen BB, Wanders SS, Borger JJ, Öllers MM, Dingemans A-MCA, et al. Mature results of a phase II trial on individualised accelerated radiotherapy based on normal tissue constraints in concurrent chemo-radiation for stage III non-small cell lung cancer. *Eur J Cancer* 2012;48:2339–46.
- [19] Turrisi AT, Kim K, Blum R, Sause WT, Livingston RB, Komaki R, et al. Twice-daily compared with once-daily thoracic radiotherapy in limited small-cell lung cancer treated concurrently with cisplatin and etoposide. *N Engl J Med* 1999;340:265–71.
- [20] De Ruyscher D, Faivre-Finn C, Nestle U, Hurkmans CW, Le Péchoux C, Price A, et al. European organisation for research and treatment of cancer recommendations for planning and delivery of high-dose, high-precision radiotherapy for lung cancer. *J Clin Oncol* 2010;28:5301–10.
- [21] Marks LB, Bentzen SM, Deasy JO, Kong F-MS, Bradley JD, Vogelius IS, et al. Radiation dose–volume effects in the lung. *Int J Radiat Oncol Biol Phys* 2010;76:S70–6.
- [22] Cox DR. Note on grouping. *J Am Stat Assoc* 1957;52:543–7.
- [23] Viganò A, Dorgan M, Bruera E, Suarez Almazor ME. The relative accuracy of the clinical estimation of the duration of life for patients with end of life cancer. *Cancer* 2000;86:170–6.
- [24] Stone P, Lund S. Predicting prognosis in patients with advanced cancer. *Ann Oncol* 2007;18:971–6.
- [25] Stockler MR, Tattersall MHN, Boyer MJ, Clarke SJ, Beale PJ, Simes RJ. Disarming the guarded prognosis: predicting survival in newly referred patients with incurable cancer. *Br J Cancer* 2006;94:208–12.
- [26] Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ* 2000;320:469–72.
- [27] Dehing-Oberije C, De Ruyscher D, van der Weide H, Hochstenbag M, Bootsma G, Geraedts W, et al. Tumor volume combined with number of positive lymph node stations is a more important prognostic factor than TNM stage for survival of non-small-cell lung cancer patients treated with (chemo)radiotherapy. *Radiat Oncol Biol* 2008;70:1039–44.
- [28] Ball D, Mitchell A, Giroux D, Rami-Porta R. IASLC Staging Committee and Participating Institutions. Effect of tumor size on prognosis in patients treated with radical radiotherapy or chemoradiotherapy for non-small cell lung cancer. An analysis of the staging project database of the International Association for the Study of Lung Cancer. *J Thorac Oncol* 2013;8:315–21.
- [29] Soliman M, Yaromina A, Appold S, Zips D, Reiffenstahl C, Schreiber A, et al. GTV differentially impacts locoregional control of non-small cell lung cancer (NSCLC) after different fractionation schedules: subgroup analysis of the prospective randomized CHARTWEL trial. *Radiother Oncol* 2013;106:299–304.
- [30] Werner-Wasik M, Swann RS, Bradley J, Graham M, Emami B, Purdy J, et al. Increasing tumor volume is predictive of poor overall and progression-free survival: secondary analysis of the Radiation Therapy Oncology Group 93–11 phase I–II radiation dose-escalation study in patients with inoperable non-small-cell lung cancer. *Radiat Oncol Biol* 2008;70:385–90.
- [31] Nalbantov G, Kietselaer B, Vandecasteele K, Oberije C, Berbee M, Troost E, et al. Cardiac comorbidity is an independent risk factor for radiation-induced lung toxicity in lung cancer patients. *Radiother Oncol* 2013;109:100–6.
- [32] Tang C, Liao Z, Zhuang Y, Levy LB, Hung C, Li X, et al. Acute phase response before treatment predicts radiation esophagitis in non-small cell lung cancer. *Radiother Oncol* 2014;110:493–8.
- [33] Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. *J Eval Clin Pract* 2001;7:97–107.
- [34] Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *J Urol* 1994;151:1283–90.
- [35] Schröder FH, van der Crujisen-Koeter I, de Koning HJ, Vis AN, Hoedemaeker RF, Kranse R. Prostate cancer detection at low prostate specific antigen. *J Urol* 2000;163:806–12.
- [36] Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level ≤ 4.0 ng per milliliter. *N Engl J Med* 2004;350:2239–46.
- [37] Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, 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.
- [38] Trott K-R, Doerr W, Facoetti A, Hopewell J, Langendijk J, van Luijk P, et al. Biological mechanisms of normal tissue damage: importance for the design of NTCP models. *Radiother Oncol* 2012;105:79–85.
- [39] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6.
- [40] Bentzen SM. From cellular to high-throughput predictive assays in radiation oncology: challenges and opportunities. *Semin Radiat Oncol* 2008;18:75–88.
- [41] Starman MHW, Lieuws NG, Span PN, Haider S, Dubois L, Nguyen F, et al. Independent and functional validation of a multi-tumour-type proliferation signature. *Br J Cancer* 2012;107:508–15.