Key challenges in normal tissue complication probability model development and validation

Van den Bosch, Lisa; Schuit, Ewoud; van der Laan, Hans Paul; Reitsma, Johannes B; Moons, Karel G M; Steenbakkers, Roel J H M; Hoebers, Frank J P; Langendijk, Johannes A; van der Schaaf, Arjen

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2020.04.012

Publication date:
2020

Radiotherapy and Oncology 148 (2020) 151–156

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Technical Note

Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy

Lisa Van den Bosch [a,*], Ewoud Schuit [b], Hans Paul van der Laan [a], Johannes B. Reitsma [b], Karel G.M. Moons [b], Roel J.H.M. Steenbakkers [a], Frank J.P. Hoebers [c], Johannes A. Langendijk [a], Arjen van der Schaaf [a]

[a] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, The Netherlands; [b] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, The Netherlands; [c] Department of Radiation Oncology (MAASTRO Clinic), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, The Netherlands

ARTICLE INFO

Article history:
Received 18 July 2019
Received in revised form 8 April 2020
Accepted 8 April 2020
Available online 23 April 2020

Keywords:
Normal Tissue Complication Probability models
Normal tissue
Statistical methods
Head and neck

ABSTRACT

Normal Tissue Complication Probability (NTCP) models can be used for treatment plan optimisation and patient selection for emerging treatment techniques. We discuss and suggest methodological approaches to address key challenges in NTCP model development and validation, including: missing data, non-linear response relationships, multicollinearity between predictors, overfitting, generalisability and the prediction of multiple complication grades at multiple time points. The methodological approaches chosen are aimed to improve the accuracy, transparency and robustness of future NTCP-models. We demonstrate our methodological approaches using clinical data.

Normal Tissue Complication Probability (NTCP) models are prediction models used in the field of radiotherapy to estimate the risk of radiation-induced complications. These models aim to translate radiation dose distributions, in combination with patient, disease and treatment characteristics, into a predicted probability that a complication will occur. In recent years, NTCP-models have been increasingly integrated into daily clinical practice. They can assist clinicians when evaluating and choosing the optimal treatment plan among various conventional and emerging techniques, such as proton therapy, by comparing the predicted complication risk of each treatment plan [1,2]. Additionally, they can be used during treatment plan optimisation to actively guide the dose distribution to lower the complication risk [3]. Obviously, this requires reliable and high quality NTCP-models. This means the models should: (1) have adequate predictive performance in terms of calibration, the agreement between predicted probabilities and observed outcome (i.e. complication) frequencies, and in terms of discrimination, the ability of the model to distinguish between patients with and without the outcome; and (2) accurately describe the effects of radiation dose to normal tissues and the risk of complications, rather than reflect a mere correlation between radiation dose and complications [4]. This is a challenging combination of two worlds: the world of prediction research and the world of causal inference. On the one side, the focus is in obtaining accurate absolute probabilities that can be used in (shared) decision making to decide on the optimal treatment technique for each individual patient. On the other side, the focus is on selecting the most relevant predictors and obtaining reliable and interpretable predictor effects, which is essential for effective treatment plan optimisation.

Accordingly, several challenges that are frequently encountered in NTCP modelling should be addressed, including missing data, non-linear response relationships, multicollinearity between predictors, overfitting, generalisability and prediction of multiple complication grades at multiple time points. We discuss and suggest methodological approaches to address these key challenges, resulting in a strategy that is designed to improve the accuracy, transparency and robustness of NTCP-models. During the development of this strategy, interpretability and clinical usability were important drivers in the decision-making process, ensuring a pragmatic approach. A schematic overview of the proposed strategy is shown in Fig. 1. A detailed explanation, including an example illustrating all aspects of the strategy in depth as well as the corresponding R code, can be found in the supplementary data. We will frequently refer to different sections of the detailed example

* Corresponding author at: University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, PO Box 30001, 9700 RB Groningen, The Netherlands.
E-mail address: l.g.l.j.van.den.bosch@umcg.nl (L. Van den Bosch).

https://doi.org/10.1016/j.radonc.2020.04.012
0167-8140/

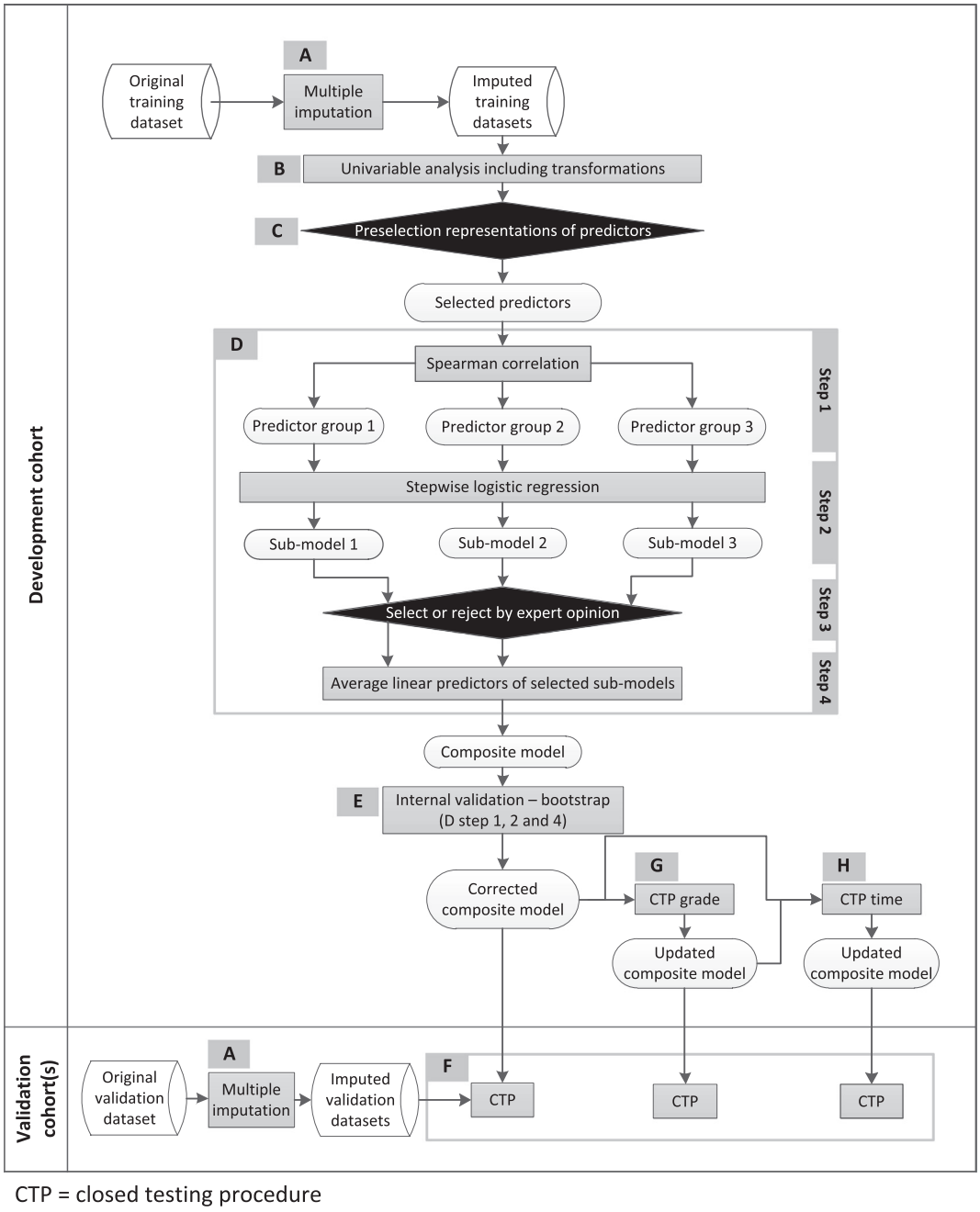

Fig. 1. Flowchart with model development and validation steps.

for further information. This will be indicated by 'example', followed by the number of the section.

Note that our approach should not be considered as the final approach to NTCP modelling, as alternative methods exist to deal with the challenges encountered during NTCP model development and validation. This is an evolving field of research and therefore our proposed strategy should be viewed as an important step forward to address the aforementioned challenges in NTCP modelling.

Missing data

Missing data is very common in medical research. Missing values can occur in both predictor and outcome variables. In NTCP modelling, outcome variables are often missing due to various reasons including non-compliance, lost to follow up due to recurrent disease, follow up being too short or death (example 2.1). A complete case analysis is often performed when (predictor or outcome) data are missing. However, in addition to substantially reducing the sample size and thus the precision of estimates, it has widely been shown that a complete case analysis also tends to lead to biased estimates of studied predictor-outcome associations [5–9]. An approach generally acknowledged to be better, is the use of multiple imputation techniques (Fig. 1A and example 2.2) [5–9]. Multiple imputation techniques, e.g. the multivariate imputation by chained equations (MICE), use all observed information of the study patients to build so-called imputation models that estimate as good as possible the distribution of the variables that have a missing value [10]. Subsequently, from these estimated distributions, a value is drawn and replaces the missing value. This is done multiple, e.g. 10, times. Each imputation set is then analysed as planned. To account for the fact that imputed values are not

formally observed but rather estimated and to penalise the analysis for its inflated precision, the results of the imputation sets (i.e. the regression coefficients and standard errors of the predictors in the final prediction model) are pooled according to Rubin's rule [11]. The pooled standard errors reflect the uncertainty of the imputation by taking the variance of the regression coefficients within an imputation set as well as the variance across imputation sets into account.

Other, more advanced, techniques are available to deal with missing repeated outcome data, e.g. multi-state models, multilevel imputation by joint modelling or fully conditional specification, or by using multilevel analysis to predict radiation-induced complications over time as these models are well equipped to handle missing outcome data [12,13]. It remains, however, important to strive for complete and high quality data, regardless of the technique used to deal with missing data, as it will not compensate for low quality data.

Non-linear transformations and representations of predictors

In NTCP modelling, all relevant candidate predictors should be considered for multivariable analysis. This includes baseline complaints, patient-, tumour- and treatment characteristics as well as dose predictors. For continuous predictors, a linear relationship with the risk of the outcome is often assumed. However, non-linear transformations, e.g. a square root or log transformation, might better describe the associations of continuous predictors. Non-linear associations are rarely included in prediction models, including in NTCP-models, even though these may substantially improve the prediction accuracy. Not exploring non-linear relationships may even preclude important predictors in the model.

Therefore, we propose to always evaluate non-linear transformations for all continuous predictors (Fig. 1B and example 3.2) [14]. A calibration plot, showing the agreement between predicted and observed outcomes, can help in visualising the need for (and type of) transformation. Also the biological and clinical rationale and plausability of transformations should be taken into consideration.

Ideally, relevant predictors are selected based on existing literature and clinical reasoning, with in the case of radiation therapy, a focus on one DVH parameter (e.g. mean, median, D1% (i.e. the dose received by at least 1% of the volume), V20 (i.e. the volume receiving at least 20 Gy)) for each organ at risk (OAR). Then, in a consecutive step, the (non-)linearity of the continuous candidate predictors can be assessed. However, if no convincing evidence exist on relevant DVH parameters, multiple DVH parameters of the same OAR may be considered potentially relevant and each of them can be transformed in multiple ways (e.g. log, square root). Furthermore, some OAR can be divided into sub-structures (e.g. ipsilateral and contralateral parts), further increasing the availability of DVH parameters since DVH parameters for both the whole organ as well as its sub-structures are available, and each of them can be transformed in multiple ways.

Transforming (multiple) DVH parameters of the same OAR (in multiple ways), however, leads to multiple representations of dose predictors of the same OAR, while data limitations often demand a reduction in the number of predictors. Therefore, we suggest a preselection of these representations of dose predictors of the same OAR before entering the predictors in the multivariable analysis (Fig. 1C and example 3.3). We use the Bayesian Information Criterion (BIC) value for this purpose, with additional penalties for non-linear transformations, organ sub-structures and unfavoured DVH parameters. These penalties are arbitrary but reflect our preference of general vs. more specific dose predictors, to facilitate the use of the NTCP-models in clinical practice and encourage acceptance by physicians. It is difficult to clearly define 'general'. But it should be regarded as the most prevailing DVH parameter for the OAR and the endpoint evaluated (e.g. mean dose, or near-maximum dose for serial organs), in its most original (i.e. not transformed) form. It is up to the researchers to decide on the penalties for transformations and/or unfavoured DVH parameters to balance evidence from the data with prior knowledge or experience. Apart from the preselection of representations of (transformed) predictor forms with similar meaning, we do not pre-select predictors based on univariable significance since this is ill-advised, as the correlation between predictors may influence their association with the outcome (stronger or weaker) once combined in a multivariable model [15].

Multicollinearity

Another challenge in NTCP modelling is multicollinearity between predictors, i.e. when two or more predictors in a multivariable model are highly correlated. Multicollinearity tends to disturb the predictor selection process. It yields an unreliable and unstable estimation of predictors' regression coefficients with stepwise logistic regression, a commonly used method in NTCP modelling (example 4.1) [16].

To deal with multicollinearity and preserve the most relevant radiation dose parameters in the models, we suggest to modify the commonly used stepwise logistic regression method in four steps (Fig. 1D and example 4.2). First, before performing stepwise logistic regression, highly correlated predictors are identified and separated into different predictor groups. Each predictor group is as large as possible without containing predictors with a mutual Spearman correlation above a specified threshold, here 0.8. This correlation threshold corresponds with a Variance Inflation Factor (VIF) of 5 or higher. A VIF $\geq$5 is often considered a cut-off value for high multicollinearity [17,18]. Second, with each predictor group, a prediction sub-model is developed using stepwise logistic regression. This is performed in each imputation set. The final predictor selection is based on the majority rule, i.e. a predictor should be selected in more than half of the imputation sets. The final sub-model is fitted on each imputation set and results are pooled according to Rubin's rules [11]. Third, sub-models are selected or rejected by model performance (BIC) and expert opinion based on clinical plausability and relevance. Finally, the selected sub-models are combined into a single logistic regression model, also called a composite model, by taking the average of the linear predictors.

With this modified stepwise logistic regression method, we aim to select the most relevant predictors, even those with high mutual correlation. Additionally, it allows for a reliable estimation of the predictor effect. Furthermore, it results in an NTCP-model that is interpretable and easy to implement in clinical practice. Other approaches, such as combining collinear predictors, using penalised regression techniques (e.g. least absolute shrinkage and selection operator or ridge regression), principal component analysis, or more advanced deep learning approaches, can also adequately deal with multicollinearity. However, with these methods reliable estimation of the predictor effect, interpretability and clinical usability of the models are not guaranteed.

Validation

Model validation is an important aspect of prediction modelling research and can be differentiated as: (1) internal validation, to correct the model for overfitting and improve the model's performance with new patients; and (2) external validation, evaluating the generalisability of the model, i.e. the variation in its perfor-

mance between populations, by applying it to a new population set and assessing whether a model adjustment (i.e. model update) is necessary to achieve appropriate performance.

For internal validation (example 5.1), we suggest the use of a bootstrapping procedure, in which model development (Fig. 1D step 1, 2 and 4) is repeated in a sufficiently high number of bootstrap samples (at least 100) that are of equal size as the development sample, but drawn at random with replacement (Fig. 1E) [19,20]. Performance of the bootstrap model is assessed in both the bootstrap sample and development cohort in terms of discrimination and calibration. Discrimination is quantified with the c-index. Calibration is quantified by the calibration intercept and slope, and assessed graphically by a calibration plot. The estimated optimism of the model is the mean difference in performance measures between the bootstrap model as applied in the bootstrap sample vs. application in the development cohort. This estimated optimism is used to shrink the performance measures and regression coefficients, thereby correcting the model for overfitting [19]. Since the univariable analysis with pre-selection of representations of predictors is not included in the bootstrapping procedure, a small portion of optimism will not be corrected for. However, the optimism induced by this pre-selection step is expected to only be minor since the univariable predictor-outcome association is not considered for selection.

For external validation (example 5.2), the model performance is evaluated for a patient cohort that was not used for model development. There is no clear definition on what qualifies as an 'externally valid model', as this also depends on the purpose of the model [21]. At least the model performance in the external validation set should be reported to evaluate the generalisability of the model. Model adjustments, such as recalibration-in-the-large (re-estimation of model intercept), recalibration (re-estimation of intercept and slope) or model revision (re-estimation of all coefficients), can be applied to improve model performance. For this, we suggest the use of a closed testing procedure; an automated method to evaluate whether and to what degree a model adjustment is needed (Fig. 1F) [22]. In case of a model revision, an external validation of the updated model is advised to evaluate the model's generalisability. When only an update of the model intercept is indicated, to adjust for differences in the prevalence of the outcome across the development and validation population, the consensus is that no additional external validation study is needed and that the model may be carefully applied in new patients, unless these new patients substantially differ from those that were used to develop and update the model [23]. For recalibration there is currently no consensus on the need of an additional external validation.

Multidimensional toxicity risk prediction

NTCP-models are generally developed for a single complication grade at a single time point (e.g. moderate-to-severe patient reported xerostomia at 12 months after treatment). Although this provides valuable information, it lacks important guidance for effective treatment optimisation as dose–response relationships might change for different complication grades (e.g. severe patient reported xerostomia), or at other time points during or after radiotherapy. Therefore, we aim to evolve to the prediction of NTCP-profiles, i.e. to predict the risk of various complication grades at multiple time points. At the same time, to facilitate usability and clinical acceptance of the NTCP-models, we prefer the model predictors to be consistent among the respective NTCP-models.

Therefore, we first suggest to develop an NTCP-model at one specific time point and for the lowest complication grade (Fig. 2, black boxes). Subsequently, a closed testing procedure can be used to evaluate this model for a higher complication grade as required (Fig. 1G, Fig. 2 light grey boxes and example 6.1), and adjacent time points (Fig. 1H, Fig. 2 white boxes and example 6.2), and adjust the model if required. In this way, model predictors are consistent for different complication grades and over time, while the response relationships may differ per grade or time point, if evident from the data. Only when a model with different predictors for another complication grade (Fig. 2, dark grey boxes) or time point is substantially better in terms of performance and clinical relevance, a model with different predictors might be preferred over a model with consistent predictors.

Different complication grades could also be modelled using an ordinal modelling approach, in which for each complication grade a different intercept will be estimated, while other response relationships remain consistent. However, we believe that for different complication grades the predictor effects might be different (e.g. stronger or weaker). Therefore, we prefer to use a closed testing procedure which also allows response relationships to change. For the modelling of different time points, a time-to-event analysis (e.g. Cox regression) can be an alternative approach for toxicities that do not recover (e.g. hypothyroidism, cerebrovascular event, myelopathy). However, for most toxicities a dynamic pattern exists, which cannot be encompassed with a time-to-event analysis.

Discussion

We discuss and suggest methodological approaches for model development and validation to address key challenges in NTCP modelling. The underlying aim was to improve the accuracy, transparency and robustness of future NTCP-models that can be used for treatment plan optimisation, treatment plan comparison, and selection of patients for emerging treatment techniques, such as proton therapy.

There are multiple ways of handling the key challenges addressed in this paper. The proposed pragmatic strategy is only one way of handling these challenges and was developed as an answer to the problems we encountered while modelling various complications. Interpretability and clinical usability of the result-

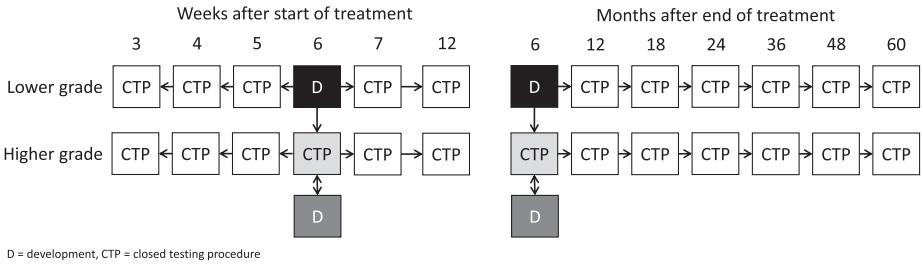

Fig. 2. Multidimensional NTCP prediction.

ing NTCP-models were key drivers in the realisation of the proposed strategy. We believe this strategy offers a solid modelling framework that may improve modelling consistency. However, careful thinking about every analysis step upfront and critical evaluation of results regarding biological or clinical plausability remains important. Throughout the manuscript we highlighted alternative, sometimes more complex, methodological approaches to address key NTCP modelling challenges. Future research should be aimed at comparing competing methodologies to unravel which approach performs better under which conditions.

Even with an adequate model development strategy, many other factors influence the quality and performance of NTCP-models [24]. This includes the right study design, the use of consistent definitions of predictor and outcome variables across centres by using standardised validated scoring systems that aim to reduce interobserver variability, instruction and training of staff to properly collect data, completeness of follow-up, and the use of uniform delineation guidelines to improve OAR delineation consistency. Poor data quality, insufficient sample size (i.e. few outcome events per estimated coefficient) and inadequate model building strategies can lead to biased and sub-optimal NTCP-models. To assess the risk of bias of prediction models, the PROBAST quality assessment tool can be used [24]. Additionally, complete and transparent reporting of model results is the key to facilitate the clinical implementation of NTCP-models. Such reports should provide full regression equations (including the intercept or baseline hazard) or nomograms to calculate NTCP-values, and information on model performance (e.g. discrimination and calibration) [25].

We aim to confirm the value of this strategy with simulation studies on the one hand and clinical validation of the produced NTCP-models on the other hand. In the simulation study, we want to compare the proposed modelling strategy to other (known) modelling strategies. We especially want to evaluate the ability of the different modelling strategies to select the most relevant predictors and accurately estimate the predictor effects, with varying collinearity between predictors. In the clinical validation, the accuracy of the produced NTCP-models will be tested in patients treated with NTCP-optimised treatment plans and with prospectively collected data. Ideally, this clinical validation should be embedded in a rapid learning healthcare system, in which NTCP-models are continuously tested and, if necessary, adjusted in an ever growing patient cohort, aiming to continuously improve the NTCP-models [26,27]. However, such an iterative system requires large amounts of patient data with sufficient follow up time. Therefore, the success and progress of new NTCP-models to converge to accurate models largely depend on the accuracy of the initial models.

Obviously, the proposed strategy can be extended to other modelling methods, such as multinomial, ordinal, or Cox regression, mixed models, or Lyman–Kutcher–Burman models, and methods that account for data from multiple centres (clustered data). Also, using a weighted average of the sub-models may further improve the performance of the composite models. Other advanced modelling techniques from the fields of machine learning and artificial intelligence may have advantages in more detailed learning abilities. However, these refinements may conflict with the transparent and pragmatic character of the proposed strategy.

Conflict of Interest

The department of Radiation Oncology has research agreements with IBA and RaySearch Laboratories. J.A. Langendijk received non-financial support and other from IBA and RaySearch Laboratories and a fee from IBA for giving a presentation at a symposium and giving consultancy. This has been paid to UMCG Research B.V.

All other authors have no conflicts of interest to disclose.

Acknowledgements

This study was financially supported by a grant from the Dutch Cancer Society (KWF project RUG 2015-7899).

We would like to thank Erik Bakker from the University Medical Center Groningen, Frederik Wesseling from MAASTRO Clinic, and Oda Wijers and Kees Visscher from the Radiotherapeutic Institute Friesland for their much appreciated help in collecting patient data that was used in the clinical example.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2020.04.012.

References

[1] Langendijk JA, Lambin P, De Ruysscher 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. https://doi.org/10.1016/j.radonc.2013.05.007.

[2] Widder J, van der Schaaf A, Lambin P, Marijnen CAM, Pignol JP, Rasch CR, et al. The quest for evidence for proton therapy: model-based approach and precision medicine. Int J Radiat Oncol Biol Phys 2016;95:30–6. https://doi.org/10.1016/j.ijrobp.2015.10.004.

[3] Kierkels RGJ, Korevaar EW, Steenbakkers RJHM, Janssen T, Van'T Veld AA, Langendijk JA, et al. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. Radiother Oncol 2014;112:430–6. https://doi.org/10.1016/j.radonc.2014.08.020.

[4] Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. Br Med J 2009;338:1373–7. https://doi.org/10.1136/bmj.b604.

[5] Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol 2010;63:721–7. https://doi.org/10.1016/j.jclinepi.2009.12.008.

[6] Sterne JAC, White IR, Carlin JB, Spratt M, Kenward MG, Wood AM, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Br Med J 2009;338:157–60. https://doi.org/10.1136/bmj.b2393.

[7] Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006;59:1087–91. https://doi.org/10.1016/j.jclinepi.2006.01.014.

[8] Moons KGM, Donders ART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol 2006;59:1092–101. https://doi.org/10.1016/j.jclinepi.2006.01.009.

[9] Schafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall/CRC Press; 1997. https://doi.org/10.1016/b978-0-12-815489-2.00016-2.

[10] van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45:1–67.

[11] Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987.

[12] Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: Competing risks and multi-state models. Stat Med 2007;26:2389–430. https://doi.org/10.1002/sim.

[13] Grund S, Lüdtke O, Robitzsch A. Multiple imputation of missing data for multilevel models: simulations and recommendations. Organ Res Methods 2018;21:111–49. https://doi.org/10.1177/1094428117703686.

[14] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. J R Stat Soc Ser C Appl Stat 1994;43:429–67. https://doi.org/10.2307/2986270.

[15] Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907–16. https://doi.org/10.1016/0895-4356(96)00025-X.

[16] Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology 2016;6:1–20. https://doi.org/10.4172/2161-1165.1000227.

[17] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer; 2013. https://doi.org/10.2174/0929867003374372.

[18] Hair JFJ, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th ed. Pearson; 2014.

[19] Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, (René),, Vergouwe, Y,, Habbema, JDF.. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774–81.

[20] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73. https://doi.org/10.7326/M14-0698.

[21] Altman DG, Royston P. What do we mean by validating a prognostic model?. Stat Med 2000;19:453–73.

[22] Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 2016;36:4529–39. https://doi.org/10.1002/sim.7179.

[23] Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76–86. https://doi.org/10.1016/j.jclinepi.2007.04.018.

[24] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8. https://doi.org/10.7326/M18-1376.

[25] 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. https://doi.org/10.1016/j.eururo.2014.11.025.

[26] Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van A, der, Schaaf, et al. Clinical trial strategies to compare protons. Semin Radiat Oncol 2018;28:79–87. https://doi.org/10.1016/j.semradonc.2017.11.008.

[27] Lambin P, Roelofs E, Reymen B, Rios E, Buijsen J, Zegers CML, et al. Rapid Learning health care in oncology ' – An approach towards decision support systems enabling customised radiotherapy. Radiother Oncol 2013;109:159–64. https://doi.org/10.1016/j.radonc.2013.07.007.