EUR Research Information Portal

Key Aspects of Prognostic Model Development and Interpretation From a Clinical Perspective

Published in:
JAMA Otolaryngology - Head and Neck Surgery

Publication status and date:
Published: 01/02/2022

DOI (link to publisher):
10.1001/jamaoto.2021.3505

Document License/Available under:
Article 25fa Dutch Copyright Act

Citation for the published version (APA):
Hoesseini, A., van Leeuwen, N., Sewnaik, A., Steyerberg, E. W., Baatenburg de Jong, R. J., Lingsma, H. F., & Offerman, M. P. J. (2022). Key Aspects of Prognostic Model Development and Interpretation From a Clinical Perspective. *JAMA Otolaryngology - Head and Neck Surgery*, *148*(2), 180-186. https://doi.org/10.1001/jamaoto.2021.3505

Link to publication on the EUR Research Information Portal

Terms and Conditions of Use
Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy
If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

Erasmus University Rotterdam
Making Minds Matter

JAMA Otolaryngology–Head & Neck Surgery | Review

Key Aspects of Prognostic Model Development and Interpretation From a Clinical Perspective

Arta Hoesseini, MD, MSc; Nikki van Leeuwen, PhD; Aniel Sewnaik, PhD; Ewout W. Steyerberg, PhD; Robert Jan Baatenburg de Jong, PhD; Hester F. Lingsma, PhD; Marinella P. J. Offerman, PhD

IMPORTANCE Prognostication is an important aspect of clinical decision-making, but it is often challenging. Previous studies show that both patients and physicians tend to overestimate survival chances. Prediction models may assist in estimating and quantifying prognosis. However, insufficient understanding of the development, possibilities, and limitations of such models can lead to misinterpretations. Although many excellent books and comprehensive methodological articles on prognostic model research are published, they may not be accessible enough for the clinical audience. Our aim is to provide an overview on the main issues regarding prediction research for health care professionals to achieve better interpretation and increase the use of prognostic models in daily clinical practice.

OBSERVATIONS The first steps of model development include coding of predictors, model specification, and estimation. Next, we discuss the assessment of the performance of a prediction model, including discrimination and calibration aspects, followed by approaches to internal and external validation and updating. Finally, model reporting, presentation, and steps toward clinical implementation are presented.

CONCLUSIONS AND RELEVANCE After thorough consideration of the research question, data inspection, and coding of predictors, one can start with the specification of a prediction model. The number of candidate predictors should be kept limited, in view of the number of events in the data, to prevent overfitting. Calibration and discrimination are 2 aspects of model performance that complement each other and should be assessed preferably at external validation. Model development should be accompanied by qualitative research among patients and physicians to facilitate the development of a valuable tool and maximize possibilities for successful implementation. After model presentation is optimized, impact studies are required to assess the clinical value of a prediction model.

JAMA Otolaryngol Head Neck Surg. 2022;148(2):180-186. doi:10.1001/jamaoto.2021.3505
Published online December 9, 2021.

Author Affiliations: Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands (Hoesseini, Sewnaik, Baatenburg de Jong, Offerman); Center for Medical Decision Making, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands (van Leeuwen, Steyerberg, Lingsma); Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, the Netherlands (Steyerberg).

Corresponding Author: Arta Hoesseini, MD, MSc, Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Dr Molewaterplein 40, 3015 GD Rotterdam, the Netherlands (a.hoesseini@erasmusmc.nl).

Although crucial in clinical decision-making, prognostication is often challenging. Research in cancer care shows that both patients[1,2] and physicians[3-7] tend to be too optimistic about survival chances. We increasingly rely on prognostic models' outcomes, while a thorough understanding of the development, possibilities, and limitations is often limited. This can lead to misinterpretation and erroneous explanation of the models' estimates. In head and neck cancer (HNC), several prognostic models have been developed. These models predict, for example, overall or progression-free survival chances[8-14] or the risk of recurrence.[15] They can be used to inform patients and support medical decision-making. It is expected that the development of prognostic models will increase over time[16] with the shift toward data-driven medicine. The growing availability of large and rich data sets, such as electronic health records, facilitates this process.[17] Many excellent books[18-20] and comprehensive methodological articles[16,21-26] are available on prognostic model research. These may not be accessible enough for the clinical audience. Herein, we aim to provide an overview on the main issues regarding prediction research for health care professionals. We provide illustrations with clinical examples to achieve better understanding and improve the development, interpretation, and implementation of prognostic models that intend to support clinical decision-making.

Discussion

We consider 7 necessary steps to develop valid prediction models with regression analysis.[18,20] These steps are used as a framework, accompanied by clinical examples.

Step 1: Defining the Research Question and Initial Data Inspection

The first step in the development of a prognostic model is to carefully state the research question and consider the available data, predictor definitions, and the outcome of interest.[18,20] Overall survival is commonly used as the outcome of interest in prediction research, while disease-free survival is preferable when comparing 2 treatments. Overall survival is simple to measure, easy to interpret, relevant for patients, and straightforward to explain. Disease-

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

free survival often requires a smaller sample size and shorter follow-up time because the end point is earlier and more frequently reached, making it a tempting outcome measure to choose.[27] Prognostic models combine several characteristics to predict the risk of an outcome for individual patients.[26] Potential predictor variables are often patient or disease characteristics at baseline (eg, age). A frequently asked question by clinicians is whether different treatment modalities could be added to a prognostic model. For example, a patient is diagnosed with recurrent HNC, and 2 options are discussed: no treatment or chemotherapy. Is it possible to predict overall survival in both groups to guide a new patient's treatment decision? Although it is tempting to make such a comparison, this is usually invalid in observational data. Attributing differences in outcome to treatment effectiveness is flawed owing to confounding by indication and could, therefore, provide invalid estimates.[28] For example, in an observational cohort, patients with HNC treated with chemoradiation have a 20% higher overall survival than patients treated by radiation alone. This result may have been confounded by age and comorbidity: younger and healthier patients are more likely to receive chemoradiation instead of radiation. Even if we control for age and comorbidity, there are always other unknown confounders, leaving us with residual confounding. Therefore, the standard for estimating the effect of treatment is using randomized clinical trial data.[29] When treatment effectiveness is studied in a randomized clinical trial, the effect estimates from this trial can be included in a prognostic model.[30] Patients in an observational cohort may have undergone treatment. Although a prognostic model may not include treatment as a prognostic factor, an average treatment effect is assumed when estimating survival chances.

Step 2: Coding of Predictors and Imputation

The coding of predictors is the next step.[18,20] To simplify the analysis and the interpretation of the results, continuous variables are often converted into categorical variables. For example, continuous body mass index levels (calculated as weight in kilograms divided by height in meters squared) are converted into 3 categories: low (<18), normal (18-25), and high (>25). Although this is often used in medical and prediction research, categorization of continuous variables in model development should be avoided because it can result in considerable loss of power and biased estimates.[31] In addition, one should be aware that most data sets contain missing values. Standard statistical software for regression analysis deletes patients with any missing value from the analysis and uses only complete cases (CCs). A CC analysis eliminates information from participants who have information on some but not all predictors, and hence, information is lost.[20] Also, bias may occur owing to systematic differences between patients with complete values and patients with missing values.[20] As a procedure for handling missing values, multiple imputation can be applied. This implies estimating the missing data multiple times to create several complete versions of an incomplete data set. These data sets are analyzed, and the results are pooled for interpretation.[32,33] This can only be completed if the missing completely at random assumption or missing at random assumption is plausible. This means that missing data have occurred completely at random or can be predicted based on other values that are observed in the data. Multiple imputation is not valid when the missingness depends on the values that are missing, also known as *missing not at random*.[18,20,32,34-36] An example of the latter is the situation when a subgroup of patients with a very poor quality of life and severe disease has not answered a questionnaire because of their condition. Multiple imputation methods are preferred over CC analysis to deal with missingness.[34,35,37] Prediction models that have been developed based on CC analysis should be interpreted with caution because a specific selection of patients is considered. In certain cases, CC analysis is the preferred method (eg, when the data are missing not at random).

Step 3: Model Specification: Identification of Candidate Predictors and Number of Outcome Events

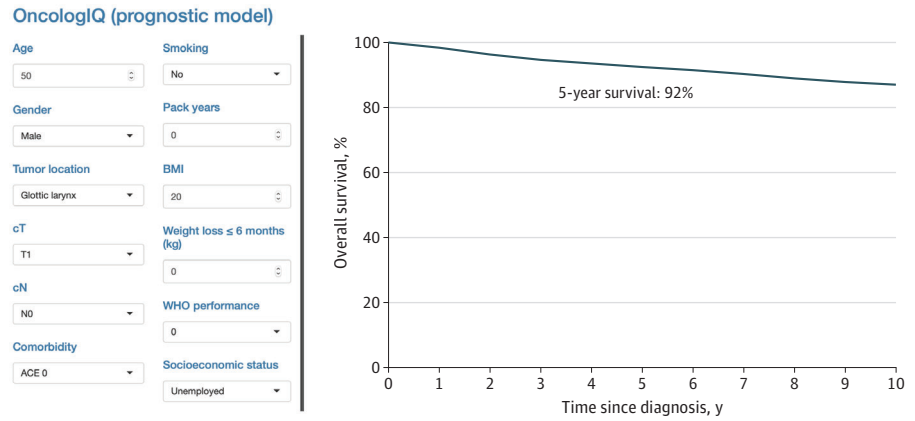
The next step is to decide which predictors should be included.[18] First, a list of candidate predictors is prespecified based on expert opinion, as well as known predictors from scientific literature. This will limit the risk of finding coincidental associations between a variable and outcome measure, also referred to as *overfitting*. This implies that the model gives valid estimations in the patient population used to develop the model but does not give valid estimations in new patients. This is often caused by testing too many predictors. Suppose that we study the predictive performance of 20 potential predictors for inclusion in a prediction model. When these variables are all noise variables (ie, variables not associated with the outcome), we will on average find 1 variable with a statistically significant ($P < .05$) effect. This variable will by chance have an extreme estimate of the regression coefficient. Also, among true predictors, those with more extreme estimates will preferentially be included. This has been labeled *testimation* bias—bias caused by first testing for statistical significance.[38] A first way to prevent testimation bias is to avoid testing too many predictors. A challenge in predictive modeling is that we often have many candidate predictors, while the sample size can be limited.[18,20] The effective sample size in prediction model research depends on the number of patients with the outcome of interest. For example, when studying a population of 1000 patients, with an outcome incidence (or *event rate*) of 2%, the effective sample size will only be 20. When considering the number of candidate predictors, a rule of thumb is that with every 10 events (eg, deaths), at most 1 continuous or 1 dichotomous candidate predictor can be included in the model.[18,20] A predictor with more than 2 categories needs more events. For example, assessing 1 predictor with 3 levels, such as education level, requires 20 events, 4 levels requires 30 events, etc. In addition, one should consider mainly predictors that are readily available in routine care. If not, this may restrain clinical implementation.[24]

Several variants of stepwise selection methods could be applied.[39] *Forward selection* starts with inclusion of the most significant candidate predictor to a model that does not contain any predictor. *Backward selection* starts with eliminating the least significant predictor from a model containing all candidate predictors.[18,20] A backward selection approach is generally preferred. Stepwise selection is based on statistical significance and might result in undue loss of clinically important predictors from the model. In some cases, nonsignificant predictors are maintained in the model for *face validity*. In other words, if there is a high belief of the importance of a specific predictor, it may therefore be maintained in the model. For example, during the update of the prognostic model OncologIQ (**Figure 1**),[40] although sex was not a statistically significant predictor, it was added to the model. Overall, it is

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

Figure 1. Prognostic Model



OncologIQ predicts the 1- to 10-year overall survival probability of patients with primary head and neck squamous cell carcinoma who are eligible for curative treatment. This model contains 12 predictors and was recently updated (C statistic = 0.71).[40] This example shows the 1- to 10-year overall survival chances of a 50-year-old man with a T1N0M0 glottic laryngeal squamous cell carcinoma (accessed November 7, 2021, at https://www.oncologiq.nl/).

important to critically assess the number of tested predictors and method of model selection in relation to the effective sample size.

Step 4: Model Estimation

In the fourth step, the model is specified and the model parameters are estimated.[18,20] The traditional approach to develop prediction models involves the use of regression models. Commonly used models are linear regression for continuous outcomes, binary logistic regression for dichotomous outcomes, and Cox regression for survival outcomes. In this article we focus on the latter. Every prediction model starts with a baseline hazard, assuming no risk factors are present. For each predictor variable in the model a regression coefficient (β) is estimated: this is an estimation of the relationship between the predictor variable and the outcome variable. A positive coefficient indicates that as the value of the predictor variable increases (eg, body mass index), the mean of the outcome variable (eg, survival) also tends to increase. Depending on your outcome measure, you can convert β in a hazard ratio (for survival outcomes) or odds ratio (for binary or dichotomous outcomes). After model estimation, shrinkage of β toward zero is one way to improve predictions for future patients.[18,20,41,42] The shrinkage factor can be used to correct for potential overfitting. The shrinkage factor is a number that is calculated in the bootstrapping procedure (step 6). The coefficients of the predictors should be multiplied by the shrinkage factor (eg, 0.8), resulting in the final shrunken coefficients. A modern estimation method to prevent overfitting is the Lasso.[20] This method combines both model selection and shrinkage of regression coefficients.

During the past decade, machine learning (ML) algorithms for prediction purposes have gained popularity in the medical field. A recent systematic review showed no evidence of superior performance of ML over logistic regression for clinical prediction modeling.[43] Because ML is highly data driven, there is a high risk for overfitting. It may only be suitable in situations with large numbers of events. It might also result in loss of clinically important predictors from the model, as some might not be recognized by these ML methods. Hence, ML is promising in pattern recognition where it has proven its added value in the field of (medical) diagnostic research.[44] In prediction research, the numbers of events are usually too small to prefer ML over regression modeling.

Step 5: Evaluation of Model Performance

Once a model is built, the quality and usefulness of the prediction model needs to be assessed. Different performance measures can be used to determine this and to quantify whether the model is useful.[18,20]

Discrimination

The discrimination of a model relates to how well a prediction model can discriminate those with the outcome from those without. The concordance index, also called *C statistic*, is the most commonly used measure to determine the discriminative performance for binary and time-to-event outcome measures (**Box 1**). The C statistic takes values between 0.5 and 1.0, where 0.5 indicates that the model is not better than chance classification and 1.0 means perfect discrimination.[23] When does a C statistic show acceptable discrimination? In other words, which cutoff is acceptable for clinical implementation of a model? There is, in fact, no specific cutoff defined in literature. To nevertheless provide some guidance, generally a C statistic below 0.6 can be considered as poor, a C statistic greater than 0.6 as moderate, a C statistic greater than 0.7 as good, and a C statistic greater than 0.8 as strong.[24] Examples of widely used prediction models in clinical practice are the Framingham risk model and the APACHE scores, which both have a C statistic of slightly higher than 0.70.[45,46] Assessing discrimination with the C statistic in the development data is the first step in looking at model performance. However, calibration and external validation are both of major importance, also because the C statistic is dependent on the heterogeneity of the population. A heterogeneous population makes it easier to discriminate between patients with and without the outcome. For example, a model with a C statistic of 0.8 can still perform poorly in a new population, either because the model was overfitted to the development data or because the new population is more homogenous.

Calibration

Prediction models provide probabilities on the chance of the event of interest. This implies uncertainty on whether the event will occur. In addition, this estimated chance is surrounded by uncertainty as well. It is often overlooked that estimated risks can be unreliable, even when the algorithms have good discrimination. We can

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

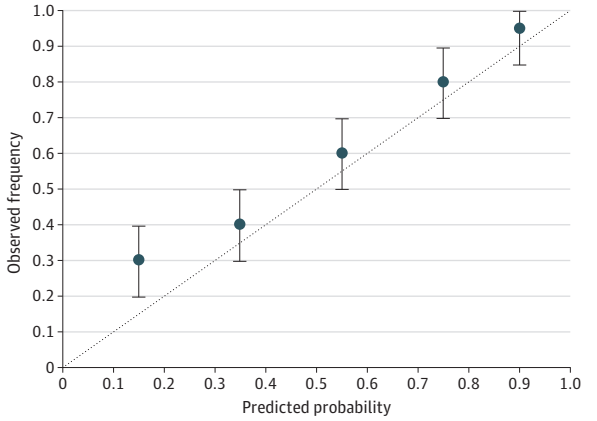
Box 1. C Statistic

When comparing pairs of patients with and without the outcome, the C statistic determines in which proportion the model gives the highest chance of the outcome to the patient with the observed outcome under study.

For example, patient X survives 5 years, and patient Y dies within 5 years. The model predicts a 5-year survival chance of 80% for patient X and 60% for patient Y. In this patient pair, the model discriminates well between the patient with and without the outcome (death). Imagine that a model has a C statistic of 0.75: this means that in 75% of all patient pairs, the model correctly identifies the patient with the outcome. So, the C statistic is the probability to be correct among pairs of patients.

assess how close the models' predictions are to the observed outcome: this is called calibration.[18,20] If we predict a 30% 5-year survival probability in a patient with an oral cavity carcinoma, the observed frequency of survival should be 30 out of 100 patients.[40] This is an example of perfect calibration. Risk estimates may be systematically too high or too low for all patients irrespective of whether they experienced the event,[47] as was the case when validating and updating OncologIQ.[40] In a calibration plot, observed outcomes are plotted against predictions from the model. Ideally, perfect predictions should be on the 45° line. In a survival context, the calibration of a model is usually assessed at a fixed time point, such as 5 years after diagnosis. In the case of OncologIQ, the predicted probabilities were systematically lower than the observed frequencies, meaning consequent underestimation of survival by the old model. **Figure 2** shows an example of this.

Figure 2. Calibration Plot



Predicted probabilities and observed frequencies in 5 quintiles. The predicted probabilities are systematically lower than the observed frequencies. Error bars indicate 95% CIs.

Step 6: Validation and Updating

Validation is an important next step.[18,20] The main goal of validation is to assess the performance of the developed model in a new data set. In other words: how well can the model predict outcomes for future patients? *Internal validation* can be used to assess the reproducibility of your model in your own data set. This is often the first step. Next, you would like to assess the validity of your model in an external data set to assess the generalizability or transportability of your model. This is called *external validation*. A key threat to validity is overfitting, often caused by testing too many candidate predictors in the model.

Internal Validation

Internal validation may be performed using bootstrapping techniques.[18,20] Bootstrapping imitates the process of sample generation from a population. Multiple samples of the same sample size are drawn with replacement from the original data set. From the bootstrapping, both an internally validated C statistic and a shrinkage factor are calculated. Another way of validating the model is using internal-external cross-validation. In this procedure, the data are split (eg, per hospital or by calendar time). The final model is developed on all available data, based on the pooled data.[48]

External Validation and Updating

Subsequently, the model performance (both calibration and discrimination) should be further tested in external data sets, to assess generalizability to new patient populations (external validation). This is a necessary step before implementation in clinical practice. When a model is potentially based on outdated data, updating methods are preferred over developing a new model. Previous studies found that recalibration methods are often sufficient to update a model, resulting in revision of the estimated predictor coefficients.[21,49,50] For example, when survival chances have improved owing to better treatment over the years, a simple way to recalibrate your model is to adjust the baseline risk or hazard. Another reason for updating is the availability of new prognostic factors. However, the added value of new predictors might be limited when specific predictors are highly associated with other factors that are already part of the prognostic model. For example, if smoking is already part of the model, the highly correlated predictor alcohol consumption is not likely to improve the models' performance. Another reason mentioned in literature for suboptimal improvement of model performance is when a specific strong predictor is common in only a few patients. For the individual patient, this prognostic information may be highly relevant; however, on a population level, this may not improve performance. For example, a rare gene mutation might be of high predictive value for an individual patient, but if the prevalence is very low then this will not improve performance of the model on population level. In addition, the discriminative performance of a specific prediction model could be lower when tested in a more homogeneous patient population because this makes it more difficult to discriminate between patients with and without the outcome.

Step 7: Model Reporting, Presentation, and Application in Clinical Practice

Previous research has shown that the quality of reporting of prediction model studies is poor.[51] To facilitate transparent reporting, the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement has been developed and consists of a 22-item checklist.[51] We advise to use the TRIPOD statement in reporting prognostic research. Next, the algorithm of the prediction model should be converted to an

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

Box 2. Summary Points

- Prognostic models can be used to inform patients and support medical decision-making.
- Generally, treatment should not be included as a prognostic factor when the model is based on observational data owing to confounding by indication.
- Dichotomization of continuous variables in model development should be avoided because it can result in considerable loss of prognostic information.
- Multiple imputation methods are preferred over complete-case analysis to deal with missing data.
- The number of candidate predictors should be limited by the number of events to prevent overfitting and testimation bias.
- After model estimation, shrinkage of regression coefficients toward zero is one way to improve predictions for future patients.
- There are several performance measures to assess the quality of a model:
  - *Discrimination:* How well can a model discriminate between patients with and without the outcome? (C statistic)
  - *Calibration:* How close are the observed outcomes to the predictions?
- Both internal and external validation of the model are of major importance before considering implementation of the model in an accessible application.
- When a model is based on potentially outdated data, updating methods are the preferred method over developing a new model.
- Model development should be accompanied by qualitative research among patients and physicians to facilitate the development of a valuable tool and successful implementation.
- After model presentation is optimized, its clinical impact should be assessed with respect to decision-making and patient outcome in an impact study.

accessible application for use in clinical practice. Nowadays, prediction models are more and more embedded in electronic patient records, making them easily accessible for health care providers. Another option is to convert a model in an online tool. An example of this is OncologIQ.[40] To facilitate successful implementation in clinical practice, such a tool should meet both health care providers' and patients' wishes. Both parties should therefore be closely involved in the development and evaluation process. A convenient way of assessing their preferences is by extensive qualitative research, such as interviews or focus groups. Not only does this facilitate the development of a valuable tool, but it may also facilitate successful implementation.[24]

Although many prognostic models are developed, only few are used in clinical practice.[24] A potential barrier of model implementation could be that the model fails to be tailored to physicians' and patients' needs, for example, when a model has been built for a specific outcome that is not relevant for patients or hard to understand, or when the tool is not easily accessible. Lack of understanding of how the model is built and concerns about its reliability could lead to lack of trust in the model. Up until now, little research has been done on the barriers and facilitators for model implementation. Various formats can be used to present models' estimations, and this depends on the targeted patient population. While some studies among patients with cancer showed that a 100-person diagram is an optimal format, others concluded that it is too confronting and preferred a pie chart.[52-54]

An important question often addressed by patients and health care providers is how the averages for a group of patients relate to a single patient. In other words: how should we interpret and communicate the models' estimates? Prognostic models aim to assess a more individual prognosis, based on several predictors. However, the estimation remains a probability calculation, while the individual will, by definition, have 1 outcome (eg, die or survive).[55] This should be taken into account during communication with patients and caregivers. We developed a clinical practice guideline to support professionals in sharing prognostic information, based on focus group research among patients with HNC.[53] This guideline includes basic steps in communicating prognostic models' estimations. We advise to show and explain all prognostic factors included in the model. This enables patients to understand which variables are used to calculate their prediction. Discuss the 5-year survival chance using a pie chart and keep it simple: speak of "survival chances" without giving further statistic details such as confidence intervals, as it remains unclear to what extent patients understand the uncertainty around prognostic models' estimates.[53,55] Until now there has been no consensus in literature about the optimal way to communicate data uncertainty.[56,57] Furthermore, health care professionals should include information on the expected quality of life.[53]

After model presentation is optimized, as a final step, one should ideally assess its clinical impact in a so-called impact study.[21,24] What is the effect of the use of the prognostic model on decision-making and patient outcome? Although this step is recommended, only few impact studies have been published.[21,24] In addition, there is no gold standard on how to evaluate the clinical impact of a prognostic model. A current trial at Erasmus MC in the Netherlands is evaluating the clinical impact of OncologIQ[40] by measuring decisional conflict[58] among patients with HNC. In this prospective clinical trial with sequential cohorts, patients who received counseling with and without OncologIQ are compared. Also, a pilot study has been completed to examine the impact of the use of OncologIQ during institutional tumor board meetings.

Conclusions

Prognostic models are increasingly important in clinical decision-making (**Box 2**). However, insufficient understanding of the development, possibilities, and limitations of such models can lead to misinterpretation. This article provides an accessible overview with clinical examples to help physicians understand the main issues regarding prediction research.

ARTICLE INFORMATION

Accepted for Publication: October 11, 2021.

Published Online: December 9, 2021. doi:10.1001/jamaoto.2021.3505

Author Contributions: Dr Hoesseini had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Hoesseini, van Leeuwen, Baatenburg de Jong, Lingsma, Offerman.
Acquisition, analysis, or interpretation of data: Hoesseini, Sewnaik, Steyerberg, Offerman.
Drafting of the manuscript: Hoesseini, van Leeuwen, Offerman.

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Hoesseini, Steyerberg.
Administrative, technical, or material support: Hoesseini, van Leeuwen.
Supervision: Sewnaik, Baatenburg de Jong, Lingsma, Offerman.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Weeks JC, Cook EF, O'Day SJ, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. *JAMA*. 1998;279(21):1709-1714. doi:10.1001/jama.279.21.1709

2. Gramling R, Fiscella K, Xing G, et al. Determinants of patient-oncologist prognostic discordance in advanced cancer. *JAMA Oncol*. 2016;2(11):1421-1426. doi:10.1001/jamaoncol.2016.1861

3. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ*. 2003;327(7408):195-198. doi:10.1136/bmj.327.7408.195

4. White N, Reid F, Harris A, Harries P, Stone P. A systematic review of predictions of survival in palliative care: how accurate are clinicians and who are the experts? *PLoS One*. 2016;11(8):e0161407. doi:10.1371/journal.pone.0161407

5. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ*. 2000;320(7233):469-472. doi:10.1136/bmj.320.7233.469

6. Chow E, Davis L, Panzarella T, et al. Accuracy of survival prediction by palliative radiation oncologists. *Int J Radiat Oncol Biol Phys*. 2005;61(3):870-873. doi:10.1016/j.ijrobp.2004.07.697

7. 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*. 1999;86(1):170-176. doi:10.1002/(SICI)1097-0142(19990701)86:1<170::AID-CNCR23>3.0.CO;2-S

8. Baatenburg de Jong RJ, Hermans J, Molenaar J, Briaire JJ, le Cessie S. Prediction of survival in patients with head and neck cancer. *Head Neck*. 2001;23(9):718-724. doi:10.1002/hed.1102

9. Datema FR, Ferrier MB, van der Schroeff MP, Baatenburg de Jong RJ. Impact of comorbidity on short-term mortality and overall survival of head and neck cancer patients. *Head Neck*. 2010;32(6):728-736.

10. Datema FR, Ferrier MB, Vergouwe Y, et al. Update and external validation of a head and neck cancer prognostic model. *Head Neck*. 2013;35(9):1232-1237. doi:10.1002/hed.23117

11. Egelmeer AG, Velazquez ER, de Jong JM, et al. Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiother Oncol*. 2011;100(1):108-115. doi:10.1016/j.radonc.2011.06.023

12. Fakhry C, Zhang Q, Nguyen-Tân PF, et al. Development and validation of nomograms predictive of overall and progression-free survival in patients with oropharyngeal cancer. *J Clin Oncol*. 2017;35(36):4057-4065. doi:10.1200/JCO.2016.72.0748

13. Larsen CG, Jensen DH, Carlander AF, et al. Novel nomograms for survival and progression in HPV+ and HPV- oropharyngeal cancer: a population-based study of 1,542 consecutive patients. *Oncotarget*. 2016;7(44):71761-71772. doi:10.18632/oncotarget.12335

14. Rios Velazquez E, Hoebers F, Aerts HJ, et al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging. *Radiother Oncol*. 2014;113(3):324-330. doi:10.1016/j.radonc.2014.09.005

15. Zhu X, Heng Y, Zhou L, Tao L, Zhang M. A prognostic nomogram for predicting risk of recurrence in laryngeal squamous cell carcinoma patients after tumor resection to assist decision making for postoperative adjuvant treatment. *J Surg Oncol*. 2019;120(4):698-706. doi:10.1002/jso.25614

16. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1-W33. doi:10.7326/M18-1377

17. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317-1318. doi:10.1001/jama.2017.18391

18. Steyerberg EW. *Clinical Prediction Models*. Springer; 2009. doi:10.1007/978-0-387-77244-8

19. Harrell Jr FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer; 2015. doi:10.1007/978-3-319-19425-7

20. Steyerberg EW. *Clinical Prediction Models*. Springer; 2019. doi:10.1007/978-3-030-16399-0

21. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606. doi:10.1136/bmj.b606

22. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247

23. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246

24. Steyerberg EW, Moons KG, van der Windt DA, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. doi:10.1371/journal.pmed.1001381

25. Hemingway H, Croft P, Perel P, et al; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595. doi:10.1136/bmj.e5595

26. Riley RD, Hayden JA, Steyerberg EW, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380. doi:10.1371/journal.pmed.1001380

27. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol*. 2005;23(34):8664-8670. doi:10.1200/JCO.2005.01.6071

28. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63(1):64-74. doi:10.1016/j.jclinepi.2009.03.001

29. Bowling A. *Research Methods in Health: Investigating Health and Health Services*. McGraw-Hill; 2014.

30. Venema E, Mulder MJHL, Roozenbeek B, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. *BMJ*. 2017;357:j1710. doi:10.1136/bmj.j1710

31. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127-141. doi:10.1002/sim.2331

32. van Ginkel JR, Kroonenberg PM. Analysis of variance of multiply imputed data. *Multivariate Behav Res*. 2014;49(1):78-91. doi:10.1080/00273171.2013.855890

33. Rubin DB. *Multiple Imputation for Survey Nonresponse*. Wiley; 1987. doi:10.1002/9780470316696

34. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59(10):1087-1091. doi:10.1016/j.jclinepi.2006.01.014

35. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. 2006;59(10):1102-1109. doi:10.1016/j.jclinepi.2006.01.015

36. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592. doi:10.1093/biomet/63.3.581

37. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models. *J Clin Epidemiol*. 2007;60(9):979. doi:10.1016/j.jclinepi.2007.03.003

38. Steyerberg EW, Van Calster B. Redefining significance and reproducibility for medical research: a plea for higher *P*-value thresholds for diagnostic and prognostic models. *Eur J Clin Invest*. 2020;50(5):e13229. doi:10.1111/eci.13229

39. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B; Collaborators. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol*. 2018;98:133-143. doi:10.1016/j.jclinepi.2017.11.013

40. Hoesseini A, van Leeuwen N, Offerman MPJ, et al. Predicting survival in head and neck cancer: external validation and update of the prognostic model OncologIQ in 2189 patients. *Head Neck*. 2021;43(8):2445-2456. doi:10.1002/hed.26716

41. Copas JB. Regression, prediction and shrinkage. *J R Stat Soc Series B Stat Methodol*. 1983;45(3):311-335.

42. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med*. 1990;9(11):1303-1325. doi:10.1002/sim.4780091109

43. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022

44. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *Eur Radiol*. 2020;30(6):3576-3584. doi:10.1007/s00330-020-06672-5

45. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619-1636. doi:10.1378/chest.100.6.1619

46. Liao Y, McGee DL, Cooper RS, Sutkowski MB. How generalizable are coronary risk prediction models? comparison of Framingham and two national cohorts. *Am Heart J*. 1999;137(5):837-845. doi:10.1016/S0002-8703(99)70407-2

47. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. doi:10.1186/s12916-019-1466-7

48. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005

49. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76-86. doi:10.1016/j.jclinepi.2007.04.018

50. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844

51. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg*. 2015;102(3):148-158. doi:10.1002/bjs.9736

52. Davey HM, Butow PN, Armstrong BK. Cancer patients' preferences for written prognostic information provided outside the clinical context. *Br J Cancer*. 2003;89(8):1450-1456. doi:10.1038/sj.bjc.6601287

53. Hoesseini A, Dronkers EAC, Sewnaik A, Hardillo JAU, Baatenburg de Jong RJ, Offerman MPJ. Head and neck cancer patients' preferences for individualized prognostic information: a focus group study. *BMC Cancer*. 2020;20(1):399. doi:10.1186/s12885-020-6554-8

54. Lagarde SM, Franssen SJ, van Werven JR, et al. Patient preferences for the disclosure of prognosis after esophagectomy for cancer with curative intent. *Ann Surg Oncol*. 2008;15(11):3289-3298. doi:10.1245/s10434-008-0068-y

55. Kattan MW. Doc, what are my chances? a conversation about prognostic uncertainty. *Eur Urol*. 2011;59(2):224. doi:10.1016/j.eururo.2010.10.041

56. Engelhardt EG, Pieterse AH, Han PK, et al. Disclosing the uncertainty associated with prognostic estimates in breast cancer. *Med Decis Making*. 2017;37(3):179-192. doi:10.1177/0272989X16670639

57. Politi MC, Han PK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. *Med Decis Making*. 2007;27(5):681-695. doi:10.1177/0272989X07307270

58. O'Connor AM. *User Manual—Decisional Conflict Scale*. Ottawa Hospital Research Institute;1993.

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Erasmus MC - Univ of Rotterdam User on 07/14/2022