Clinical prediction models: diagnosis versus prognosis

Maarten van Smeden [a,*], Johannes B Reitsma [a], Richard D Riley [b], Gary S Collins [c,d], Karel GM Moons [a]

[a] *Julius Center for Health Science and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands*
[b] *Centre for Prognosis Research, School of Medicine, Keele University, Staffordshire, UK*
[c] *Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK*
[d] *NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK*

Abstract

Clinical prediction models play an increasingly important role in contemporary clinical care, by informing healthcare professionals, patients and their relatives about outcome risks, with the aim to facilitate (shared) medical decision making and improve health outcomes. Diagnostic prediction models aim to calculate an individual's risk that a disease is already present, whilst prognostic prediction models aim to calculate the risk of particular heath states occurring in the future. This article serves as a primer for diagnostic and prognostic clinical prediction models, by discussing the basic terminology, some of the inherent challenges, and the need for validation of predictive performance and the evaluation of impact of these models in clinical care. © 2021 Elsevier Inc. All rights reserved.

Keywords: Diagnostic; Prognostic; Prediction models; Model performance; Model impact; Validation; Reporting guidelines

Clinical prediction models: diagnosis versus prognosis

1. Introduction

Clinical prediction models usually fall within one of two major categories: diagnostic prediction models that estimate an individual's probability of a specific health condition (often a disease) being currently present, and prognostic prediction models that estimate the probability of developing a specific health outcome over a specific time period [1–3]. A clinical prediction model is usually developed to estimate probabilities, often simply called *risk*, for individual patients with the aim to inform them, their relatives and healthcare providers about diagnosis or prognosis, to help with making better (shared) testing and treatment decisions, or for making risk-stratifications for therapeutic trials [2].

An example of a diagnostic prediction model is the *Wells rule* to aid the diagnosis of deep venous thrombosis (DVT) which has shown to successfully decrease the number of unnecessary referrals of low-risk primary care patients suspected of DVT to secondary care [4]. Besides triage and referral decisions, a diagnostic prediction model may also be used to identify patients at an increased risk for the presence of the target condition to undergo expensive, time consuming or invasive reference standard procedures such as biopsies or tests that require advanced imaging techniques. The *EuroSCORE II* is an example from the large pool of available long-term cardiovascular prognostic prediction models [5] that can play an important role in preventive treatment decisions and lifestyle changes. Prognostic prediction models exist for virtually all first-time and recurrent major health conditions.

2. Definitions and differences

The key difference between diagnostic and prognostic prediction models is in the temporal relationship between the moment of prediction and the outcome of interest (Fig. 1). In a diagnostic prediction model, the outcome of interest is the current health condition of the patient at the moment of prediction. The natural study design for diagnostic prediction models is a cross-sectional study, in which data are collected from a group of patients suspected of having the *target condition*, and measures both the outcome (target condition status: presence versus absence) and *predictors* at the same moment in time or with very little time in between. An exception is the explicit use of follow-up to observe whether disease occurs in subjects

* Corresponding author. Tel: +31 88 75 509 40.
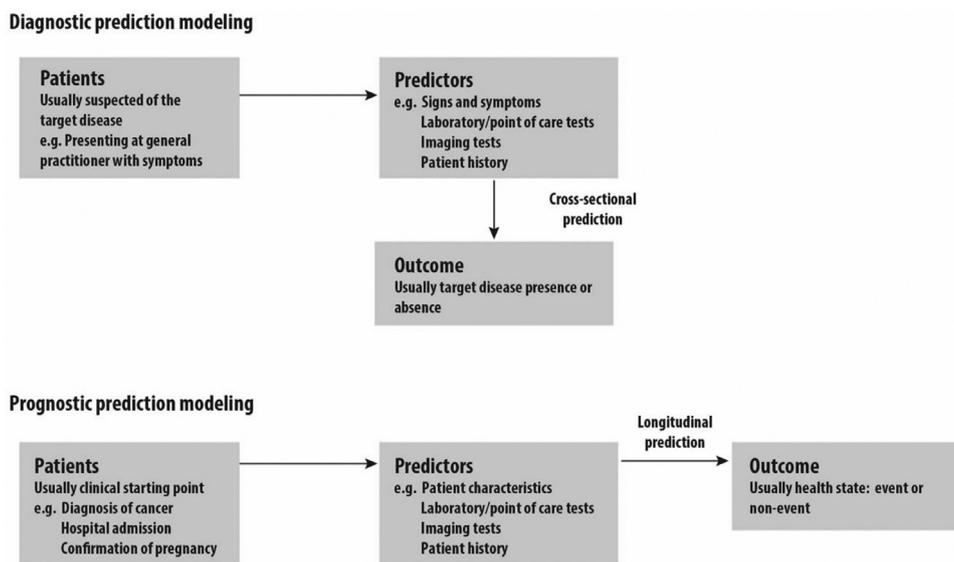 *E-mail address:* m.vansmeden@umcutrecht.nl (M. van Smeden).

Fig. 1. Schematic representation of diagnostic and prognostic model studies, adapted from TRIPOD reporting guideline [1].

in whom the reference standard cannot be performed, e.g., in imaging studies where it is impossible to perform a biopsy in a subject in which no lesion is detected during imaging (relies on the assumption that detection of target condition at follow-up accurately reflects the target condition at moment of prediction).

For prognostic prediction models, the focus is on predicting a future health outcome that occurs after the moment of prediction, also using predictors available at the moment of prediction. The *prediction horizon* – how far ahead in time the model aims to predict outcome occurrence by – needs to be established. In a prognostic model study, the time patients can be followed-up to register the outcome should match this prediction horizon. Individuals for whom the outcome has not been recorded before the end of the prediction horizon, because they have not (yet) experienced the outcome, or have dropped out of the study (*loss-to-follow up*) are *censored,* which can be accounted for using survival modeling methods.

Likewise, diagnostic models, which are commonly developed using logistic regression modeling or some variation thereof, may also suffer from incomplete outcome measurement (*partial verification*), for instance because reference standards are invasive or expensive (indeed, as mentioned above a reason why the diagnostic model may be developed in the first place). The impact of such partially observed or missing data on the results and approaches to circumvent their detrimental effects are beyond the scope of this text; information and references can be found in [1].

3. Prediction model development, performance and impact

The development of a diagnostic or prognostic model requires several common analysis steps and decisions to be taken by the modeler. In brief, after the targets of prediction are determined (e.g., outcome, target population and intended moment of use) and the dataset is prepared, building the prediction model requires decisions regarding the modeling framework (e.g., logistic regression), the candidate predictors to examine, how to code the predictors and determine the functional form of the relationship between the predictors and outcome (e.g., a nonlinear effect of a continuous predictor, such as patient age, using a spline function), dealing with missing data in predictor and outcome, and possibly selection among candidate predictors (e.g., using backward elimination in a regression model). A more detailed overview for developing (and validating) a clinical prediction model is beyond the scope of this text and found elsewhere [11].

Another shared characteristic of diagnostic and prognostic prediction models is that the focus is not on the individual predictor effects, but rather the performance of the model as a whole (sometimes the interest is in the ability of a new predictor to improve the model as a whole). The statistical performance of a clinical prediction model is typically evaluated in terms of its *discrimination* and *calibration.* Discrimination is the extent to which predicted risks differentiate between individuals with and without the outcome, and is typically quantified by calculating the C-index [2]. *Calibration* is the extent to which predicted risks correspond to proportions of observed outcomes, and is often evaluated by calculating (for example) the calibration slope, expected:observed ratios, and calibration plots. Calibration is often neglected but is critical [6], especially when identifying thresholds of risk (e.g. using decision curves) for clinical decision making [7].

Evaluation of the performance of a diagnostic or prognostic prediction model on the same data that were used in its development will lead to results that are too optimistic [1]. During model development, a clinical prediction model

therefore requires thorough *internal validation* to estimate and correct for this optimism, using bootstrap or cross-validation techniques. Usually external validation, using data from a different population that were not used for development, should then be carried out before the model should be considered for use in practice. To avoid disappointment about the model performance following validation, it is essential that the development dataset has a sufficiently large sample size relative to the complexity of the model [8], the data are high quality with an unbiased sampling of participants (e.g., consecutive) and the analysis process avoids common harmful modeling choices [9].

Finally, it is worth noting that development and validation of a diagnostic or prognostic model are only the first steps toward implementation. As clinical prediction models are expected to vary in their predictive performance over time and place, they require repeated evaluation over time (*temporal validation*) and place, when applied in different health care settings (*geographic validation*) and, if necessary, the model may need revising (*recalibrating* or *updating*). However, the ultimate goal of a clinical prediction model goes beyond predictive performance – the goal is to improve medical decision making for which good predictive performance is only a prerequisite but far from a guarantee. The evaluation of the ability of a model to improve medical care and to ensure no disproportionate harm is inflicted, requires an evaluation of relevant patient outcomes in *model impact studies* [10], which at the point of this writing are still too rare.

4. Pointers

- Common pitfalls in diagnostic and prognostic model studies include: incomplete or unclear reporting, use of nonrepresentative data, no internal or external prediction validation, modeling and validation with too low sample size, dichotomization of predictors, data-driven selection of predictors in small sample size settings, and missing data that are ignored.
- For reporting guidance, risk of bias assessments and checklists for diagnostic and prognostic model studies: see TRIPOD [1,2] and PROBAST [3].
- Developing new clinical prediction models should be avoided if there are relevant existing prediction models already available for the same outcome or target population that can be validated, updated, extended or evaluated for their impact.

Author statement

Maarten van Smeden: Conceptualization, writing – original draft, visualization, project administration; **Johannes B Reitsma:** Conceptualization, writing – review and editing; **Richard D Riley:** Conceptualization, writing – review and editing; **Gary S Collins:** Conceptualization, writing – review and editing, **Karel GM Moons:** Conceptualization, writing – review and editing.

Declarations of competing interest

None.

References

[1] 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. Explanation and elaboration paper of TRIPOD, a reporting guideline for diagnostic and prognostic model development and validation studies. Contains many explanations and key references for further reading.

[2] 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. *The TRIPOD statement. More information about TRIPOD, the reporting guideline and reporting checklists are found on: https://www.tripod-statement.org/. Extensions of TRIPOD to reporting of prediction models for clustered data and machine learning/artificial intelligence are currently in development.*

[3] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1. Explanation and elaboration paper of PROBAST, a risk of bias tool for diagnostic and prognostic model development and validation studies. Contains many clues on how to avoid risk of bias and contains key references for further reading.

[4] Geersing GJ, Zuithoff NPA, Kearon C, Anderson DR, ten Cate-Hoek AJ, Elf JL, et al. Exclusion of deep vein thrombosis using the wells rule in clinically important subgroups: individual patient data meta-analysis. BMJ 2014;348:g1340.

[5] Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.

[6] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230. *Prediction model calibration is often ignored in the development and validation of diagnostic and prognostic prediction models. This paper explains why calibration is an essential part of the evaluation of prediction model performance.*

[7] Wynants L, van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B. Three myths about risk thresholds for prediction models. BMC Med 2019;17:192. *Describes three common myths about defining risk thresholds, e.g. a cut-off point on the probability scale above which patients are classified as a predicted case, and presents methods to evaluate and select context-dependent thresholds.*

[8] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441. Presents formulas for the minimal sample size required for developing a diagnostic or prognostic prediction model. Comes with software in R and Stata to simplify calculations.

[9] Steyerberg EW, Uno H, Ioannidis JPA, Van Calster BCollaborators. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol 2018;98:133–43. Illustration the impact of a small sample size and use of common but suboptimal statistical approaches.

[10] Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606. b606*Discusses how to determine the practical value of prediction models with focus on prognostic models*.

[11] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35:1925–31. *Describes steps to improve development and validation of clinical prediction models*.