Contemporary issues in static and dynamic prediction

Chen, Yuntao

DOI:
10.33612/diss.232289260

Chapter 7
General discussion

Summary

Prediction models that estimate the probabilities of developing a specific disease (diagnostic model) or a specific endpoint of disease (prognostic model) given a set of subject's characteristics are closely connected to personalized medicine of which the key idea is to base medical decisions on individual patient characteristics rather than on population averages. Depending on decision point, prediction models can be divided into two categories: static prediction models (making one-off decision) and dynamic prediction models (making dynamically updated decisions). While multivariable logistic and Cox regression are commonly used to develop prediction models, they are not the master key to every situation. Various issues such as clustered data, competing risks and time-dependent variable may occur when a simple logistic or Cox model cannot estimate the risk correctly in static and dynamic prediction. Although adapted or more advanced approaches have been developed to address those issues in medical statistics field, they are not appropriately applied in clinical research. To fill this gap, this thesis illustrated how sophisticated statistical models can be appropriately applied to obtain better predictions using a series of clinical case studies.

The first part of the thesis consists of two studies focusing on the issue of between-study heterogeneity and competing risks in static prediction. **Chapter 2**(1) presented a practical example of developing a clinical prediction model by addressing between-study heterogeneity in the population. The purpose of this study was to develop a

heart failure (HF) phenotype stratified model for predicting 1-year mortality in patients admitted with acute HF through an individual participant data meta-analysis of four European cohorts. Two aspects of heterogeneity were addressed in this study. One is the potential different baseline mortality rates of three HF subtypes. The other one is the potential different predictor-outcome associations across HF phenotypes. Using data collected from 3577 patients across four European cohorts, we developed an HF phenotype stratified model with 11 readily available predictors. Four of the predictors, namely systolic blood pressure, serum creatinine, myocardial infarction, and diabetes, influenced mortality risk differently in the HF phenotypes. The model showed excellent discriminative ability with a weighted internal-external cross-validated AUC of 0.79 (0.74-0.83) for HF with reduced ejection fraction, 0.74 (0.70-0.79) for HF with mid-range ejection fraction, and 0.74 (0.71-0.77) for HF with preserved ejection fraction. Calibration was also good with the average predicted 1-year mortality risks close to the average observed ones, especially after recalibration of the baseline mortality risks. Our model showed excellent generalizability across four European cohorts and may provide a useful tool in HF phenotype-specific clinical decision making. In **Chapter 3**(2), we used data from a retrospective cohort in Wuhan, China to develop a prognostic model to predict in-hospital mortality in COVID-19 patients. A competing risk analysis treating discharged alive as competing risk was conducted in this study to avoid over-estimation of in-hospital mortality. The final model with five routinely measured demographic and clinical predictors was internally validated using the bootstrap approach and externally

validated in another cohort in Wuhan. The model showed very good discriminative and calibration accuracy. The model may assist physicians to early stratify the patients according to the estimated mortality at 7-day (14-day or 30-day) after admission, thus giving patients targeted supporting care and better allocating the limited medical facilities (e.g., ventilators), especially when critical care capacities are overwhelmed.

The second part of the thesis includes three studies regarding dynamic prediction. **Chapter 4 and 5** presented the application of two different joint modelling approaches in the clinical setting, illustrating how we could model a time-dependent variable and study its association with a survival outcome. **Chapter 6** empirically evaluated the predictive accuracy of the above-introduced joint models in a clinical context. Specifically, in **Chapter 4**(3), we introduced the shared random effects model (SREM) to study the association between a continuous time-dependent variable and a survival outcome using an example in respiratory medicine.. We introduced readers to the terminology of the SREM and explained why compared to the more commonly applied time-dependent Cox model (TDCM), the SREM is expected to result in less biased estimates of the effect of the time-dependent variable. A practical example studying the association between repeated central apnea hypopnea index (cAHI) measurements and cardiovascular mortality was used to illustrate the approach. Both the SREM and the TDCM were fitted to estimate the association between the current value of cAHI and the risk of cardiovascular mortality. We found the estimated HR from the TDCM to be attenuated as expected

based on the statistical literature. Our study showed that the SREM may be a more appropriate approach to study the association between a continuous time-dependent variable and a survival outcome. **Chapter 5** presented an application of a misclassification model, an important special case of the hidden Markov model (HMM), to study the association between a categorical time-dependent variable and a survival outcome. Our aim was to study the association between chronic kidney disease (CKD) stage and new-onset HF by taking into account declining kidney function with age and potential misclassification of CKD stages. Previous studies have shown that kidney dysfunction is associated with incident HF(4). However, less is known about how a subject's risk of developing HF changes with age in response to the subject's change in CKD stage with age. To address this question, we fitted a multi-state model with three CKD stages as transient states and new-onset HF (outcome of interest) and death before new-onset HF (competing risk) as absorbing states. We found that the risk association between CKD stage and new-onset HF attenuated with age on the relative scale but strengthened with age on the absolute scale. Besides, the estimated incidence rates between states (e.g., less severe CKD stage to more severe CKD stage and various CKD stages to HF/death) were very informative in describing the individual risk of transitioning from one state to another, thus delineating the cardiorenal disease course. The SREM introduced in **Chapter 4**, and the HMM used in **Chapter 5** can be both used to dynamically predict disease prognosis. In **Chapter 6**, the predictive performance of the two models were emperically compared in the context of dynamically predicting all-cause mortality

for patients with acute HF based on serial NT-proBNP measurements. Our results demonstrated that dynamic prediction approaches incorporating serial NT-proBNP measurements have the potential to improve prediction accuracy compared to the approach using baseline NT-proBNP measurement only. The HMM performed better than the SREM in dynamically predicting mortality using serial NT-proBNP measurements in this patient cohort. Therefore, the HMM can be a useful tool to aid dynamic prediction in clinical research.

Application scenarios for dynamic prediction

Dynamic prediction is very useful to update predictions in response to changes in the disease status of patients. However, the increase in predictive performance that one would gain from dynamic prediction may vary on a case to case basis. Previous research found dynamic prediction to be very useful in predicting disease prognosis in patient populations. For example, Fontein et al.(5) developed a dynamic prediction model to predict 5-year survival in early breast cancer patients treated with ecdocrine treatment (ET). The model was able to predict additional 5-year survival at any prediction timepoint up to 3 years after starting adjuvant ET and showed a cross-validated c-index that improved from 0.70 to 0.79. This means that the model performed better in distinguishing patients who will versus those who will not die of breast cancer by incorporating additional information after diagnosis. Similarly, dynamic prediction has been shown to improve the predictive ability in other disease areas including but not limited to prostate cancer(6,7), kidney disease(8), diabetes(8)

and cystic fibrosis(9). However, dynamic prediction seems less useful in predicting disease/mortality in healthier general populations. Using data from the Atherosclerosis Risk in Communities (ARIC) study, Barrett et al. found dynamic prediction models incorporating repeated measurements of systolic blood pressure resulted in little improvement in cardiovascular risk prediction compared to static prediction models(10). A similar conclusion was reached by Paige et al. in a more-powered individual participant data meta-analysis with 191,445 adults from 38 cohorts(11). The different effects of age in primary versus tertiary prevention settings may partly explain the divergent predictive performance of dynamic prediction in the above-mentioned two scenarios. When developing dynamic prediction models in patient populations (teriary prevention), the 'survived' population some time after baseline (e.g., one year) is very different from the population at baseline in the sense that the 'survived' population either had a lower baseline risk or benefited more from the interventions provided. Age in this case cannot explain much of the heterogeneity between the population at baseline and the population after baseline. However, when developing dynamic prediction models in general populations (primary prevention), the heterogeneity between the population at baseline and the population after baseline may be largely explained by the aging effect. For example, assuming two subjects aged 60 and 61 with similar values of other risk factors for cardiovascular disease, the subject aged 60 essentially becomes the one aged 61 after surviving one year. Therefore, there is little gain to develop a dynamic prediction model compared with a static prediction model in which age is included as a predictor. Although

dynamic prediction is of more value in predicting prognosis in the patient populations, it may be still very useful for predicting diseases in general populations. For example, when certain disease has a highly predictive biomarker that varies considerably over time, dynamic prediction models may be very useful in making personalized screening strategies in high-risk populations.

Dilemma between predictive accuracy and model complexity

Numerous prediction models have been developed for all kinds of diseases, e.g., 363 models predicting cardiovascular disease (CVD)(12) and 796 models predicting outcomes for patients with CVD(13). Nevertheless, only a small fraction of those models are incorporated in clinical guidelines. One of the potential reasons behind this gap is the dilemma between predictive accuracy and model complexity. In this thesis, we were devoted to building more advanced prediction models to achieve better predictive accuracy. Beyond this thesis, medical statisticians are also working hard to improve predictive accuracy by considering many other methodological issues, such as avoiding dichotomizing continuous predictors(14)and non-linearity modelling(15). When applied and interpreted correctly, improved predictive accuracy that results from a wider adoption of more advanced prediction models can be translated into gains in clinical benefit. However, most of these advanced prediction models are also more complex and less friendly to be used in clinical practice. For example, in **Chapter 6**, we developed joint models to dynamically predict mortality in patients with HF by incorporating serial NT-proBNP

measurements. The joint models are very demanding in the sense that it requires the study to repeatedly measure NT-proBNP and that it also requires model users to well understand the model terminology. Another example of this dilemma is that machine learning approaches could not get wider clinical adoption despite achieving more accurate predictions compared with traditional multivariable regression models(16). Lack of transparency means that doctors and patients don't understand how predictions are made. In this regard, doctors cannot explain the clinical decisions to patients. Consequently, patients are reluctant to trust the clinical decisions that doctors made for them.

There are many ways to balance model accuracy and model complexity in order to make better clinical decisions. We find it may be promising to correct the bias in different stages. For example, when the objective is to correct biases from a naive approach, the common way as we did in this thesis is to use a more complex model in the stage of model development. This approach, however, requires the model developers to illuminate the complex methodology to model users. Overall, fewer efforts have been put into making the prediction model more transparent, although we do see some endeavors in this direction like the utilization of the risk-scoring system(17) and Shiny App(18). Patients, doctors and statisticians must collaborate more to develop prediction models that can be both statistically sound and clinically useful. An alternative way that could be a future direction is to correct the bias in the stage of decision-making. For example, when developing a prognostic model in the presence of competing risks, the simple Cox model can still be used to calculate the

event risk given that a bias-adjusted cut-off point is used to make medical decision. However, the quantity of bias from the simple model should be precisely estimated to obtain the correct cut-off point.

As an alternative to the dynamic prediction approaches considered in this thesis, landmarking(19) is a more transparent approach to obtain dynamic predictions. The idea of landmarking is very straightforward: construct a separate prediction model for each time period that is relevant to the study objective. For time-dependent variables, the values at the baseline of each separated model are used as time-fixed variables in these analyses. The landmarking approach can easily incorporate any type of information about the subjects' history. Also, it results in relatively parsimony models with fewer parameters compared to more complex joint models(20). For example, in the above-mentioned example that predicts HF prognosis using serial NT-proBNP measurements, only one parameter is needed to estimate the effect of NT-proBNP on mortality in each separate model in landmarking approach while far more parameters are needed in estimating the NT-proBNP trajectory in joint model. However, in the naïve landmarking approach, multiple model estimations are needed to make dynamic predictions. People may argue that some advanced versions of landmarking approach such as the landmarking supermodel proposed by van Houwelingen(19) can be used to avoid multiple model estimations. These more advanced verions go however against the transparent feature of the landmarking approach and do not match with the general idea of building clinical prediction models. In summary, the choice of using joint models, naïve

landmarking, or landmark supermodels is not only dependent on model accuracy, but also related to the degree to which model users can accept the model complexity.

Time-dependent variables in etiological studies

While the joint models that incorporate the time-dependent variables have the potential to improve predictive accuracy, great care should be exercised when modelling time-dependent variables to answer causal questions. In addition to the time-dependent exposure that is of interest, confounders may be also measured repeatedly during follow-up, resulting in the issue of time-varying confounding(21). Previous research showed that inadequately adjustment for time-varying confounding may lead to biased estimates because the time-varying confounder could be both a confounder and an mediator in the causal pathway(22,23). Another difficulty in modelling time-dependent variables in etiological studies is the results interpretation. Compared with the traditional Cox model where the regression coefficient can be interpreted as a long-term effect of a risk factor on the outcome, the models accounting for time-dependent variables, e.g., the SREM in **Chapter 4**, are typically addressing short-term effects and the regression coefficients are always a weighted average of those short-term effects. Therefore, as Dekker et al. discussed in their study, clinical reasoning and a sound research question should drive the choice of the model used to analyze the data(23). For example, it may be less useful to investigate the short-term effect of some traditional lifestyle risk factors like BMI on mortality since people should be more likely to die after long exposure to obesity.

On the contrary, it may be quite meaningful to investigate the short-term effect of biomarkers like NT-proBNP on the prognosis of HF.

Concluding remarks

This thesis focused on some practical issues in clinical prediction studies, such as between-study heterogeneity, competing risks, and time-dependent variables. Clinical case studies were used to demonstrate how advanced statistical approaches can be appropriately applied to obtain better predictions. Furthermore, we empirically evaluated different dynamic prediction approaches in a clinical context to inform the application scenarios of those complex models. Future studies are needed to balance the tradeoff of predictive accuracy and model complexity when developing clinical prediction models.

References

1. Chen Y, Voors AA, Jaarsma T, Lang CC, Sama IE, Akkerhuis KM, et al. A heart failure phenotype stratified model for predicting 1-year mortality in patients admitted with acute heart failure: results from an individual participant data meta-analysis of four prospective European cohorts. BMC Med. 2021 Dec;19(1):21.

2. Li J, Chen Y, Chen S, Wang S, Zhang D, Wang J, et al. Derivation and validation of a prognostic model for predicting in-hospital mortality in patients admitted with COVID-19 in Wuhan, China: the PLANS (platelet lymphocyte age neutrophil sex) model. BMC Infect Dis. 2020 Dec 17;20(1):959.

3. Chen Y, Postmus D, Cowie MR, Woehrle H, Wegscheider K, Simonds AK, et al. Using joint modelling to assess the association between a time-varying biomarker and a survival outcome: an illustrative example in respiratory medicine. European Respiratory Journal [Internet]. 2021 Feb 1 [cited 2021 Jul 31];57(2). Available from: https://erj.ersjournals.com/content/57/2/2003206

4. Kottgen A, Russell SD, Loehr LR, Crainiceanu CM, Rosamond WD, Chang PP, et al. Reduced Kidney Function as a Risk Factor for Incident Heart Failure: The Atherosclerosis Risk in Communities (ARIC) Study. JASN. 2007 Apr;18(4):1307–15.

5. Fontein DBY, Klinten Grand M, Nortier JWR, Seynaeve C, Meershoek-Klein Kranenbarg E, Dirix LY, et al. Dynamic prediction in breast cancer: proving feasibility in clinical practice using the TEAM trial. Ann Oncol. 2015 Jun;26(6):1254–62.

6. Yu M, Taylor JMG, Sandler HM. Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival–Cure Model. Journal of the American Statistical Association. 2008 Mar 1;103(481):178–87.

7. Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. Biostatistics. 2009 Jul;10(3):535–49.

8. Musoro JZ, Struijk GH, Geskus RB, Ten Berge I, Zwinderman AH. Dynamic prediction of recurrent events data by landmarking with application to a follow-up study of patients after kidney transplant. Stat Methods Med Res. 2018;27(3):832–45.

9. Keogh RH, Seaman SR, Barrett JK, Taylor-Robinson D, Szczesniak R. Dynamic Prediction of Survival in Cystic Fibrosis: A Landmarking Analysis Using UK Patient Registry Data. Epidemiology. 2019 Jan;30(1):29–37.

10. Barrett JK, Sweeting MJ, Wood AM. Dynamic Risk Prediction for Cardiovascular Disease: An Illustration Using the ARIC Study. In: Handbook of Statistics [Internet]. Elsevier; 2017 [cited 2022 May 9]. p. 47–65. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0169716117300044

11. Paige E, Barrett J, Pennells L, Sweeting M, Willeit P, Di Angelantonio E, et al. Use of Repeated Blood Pressure and Cholesterol Measurements to Improve Cardiovascular Disease Risk Prediction: An Individual-Participant-Data Meta-Analysis. Am J Epidemiol. 2017 Oct 15;186(8):899–907.

12. 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 May 16;353:i2416.

13. Wessler BS, Lai YH L, Kramer W, Cangelosi M, Raman G, Lutz JS, et al. Clinical Prediction Models for Cardiovascular Disease. Circulation: Cardiovascular Quality and Outcomes. 2015 Jul;8(4):368–75.

14. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006 Jan 15;25(1):127–41.

15. Nieboer D, Vergouwe Y, Roobol MJ, Ankerst DP, Kattan MW, Vickers AJ, et al. Non-linear modeling was applied thoughtfully for risk prediction: the Prostate Biopsy Collaborative group. J Clin Epidemiol. 2015 Apr;68(4):426–34.

16. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ. 2019 Mar 12;364:l886.

17. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks: Competing Risks and Risk Scores. Statist Med. 2016 Sep 30;35(22):4056–72.

18. Jalali A, Alvarez-Iglesias A, Roshan D, Newell J. Visualising statistical models using dynamic nomograms. PLOS ONE. 2019 Nov 15;14(11):e0225253.

19. Houwelingen HCV. Dynamic Prediction by Landmarking in Event History Analysis. Scand J Statist. :16.

20. van Houwelingen HC, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. Lifetime Data Anal. 2008 Dec;14(4):447–63.

21. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. BMJ. 2017 Oct 16;j4587.

22. Fisher LD, Lin DY. Time-Dependent Covariates in the Cox Proportional-Hazards Regression Model. Annual Review of Public Health. 1999;20(1):145–57.

23. Dekker FW, de Mutsert R, van Dijk PC, Zoccali C, Jager KJ. Survival analysis: time-dependent effects and time-varying risk factors. Kidney Int. 2008 Oct;74(8):994–7.