Contemporary issues in static and dynamic prediction

Chen, Yuntao

DOI:
10.33612/diss.232289260

Chapter 1
General introduction

Survival analysis in medical research

Survival analysis denotes the statistical modelling and analysis of the expected duration of time until a pre-specified event. This event could be anything, such as the occurrence of heart failure (HF) or death from HF in medical research. Analysis of survival (time-to-event) data has some special properties. First, time to event is always positive and its distribution is often skewed. For example, in a study to predict mortality for acute HF patients, mortality was much higher in the first year than in the subsequent follow-up(1). Second, follow-up information is often incomplete. The most common type of incomplete data is called right censoring and occurs when a subject drops out before the study ends or when a subject is event-free at the end of the study. For example, in the same study mentioned above, patients may stop participating in the study before they experience the event of interest or they may be still alive at the end of the study. In these cases, we only observe a lower bound for the survival time rather than the exact event time. Right censoring is common in medical research, where in most cases one cannot wait until the last participant experiences the event and drop-outs are regularly experienced.

One of the most frequently used approaches in survival analysis is the Kaplan-Meier estimator(2) and a plot of the Kaplan-Meier estimator is very informative to explore the survival of a population as a whole or to compare the survival of two or more groups. However, the Kaplan-Meier method is less useful in exploring the association of continuous variables with survival. It may be also difficult to analyze

several variables simultaneously. Another widely used approach is the Cox proportional hazards (PH) model, which is usually used in cohort studies to relate the rate of event occurrence with covariates. While the Cox PH model is a versatile model that is applicable in a wide range of cases, it does rely on certain statistical assumptions that are sometimes violated. Violation of the assumptions of the Cox model will lead to biased estimates of a subject's absolute event risk, thus hampering clinical decision making that is based on these predicted event risks.

For example, in the Rotterdam study, Wolbers et al. found that the standard Cox model overestimated the 10-year risk of coronary heart disease in older women(3). They attributed this overestimation to the increased occurrence of death in this elderly population, which are treated as censored observations in the standard Cox model rather than as competing events that preclude the event of interest (coronary heart disease) from occurring. Another setting in which the standard Cox model results in biased estimates is when the time-to-event data are clustered(4). Such data arise when each study participant can potentially experience several events (e.g., multiple infections after hospitalization) or when there exists some natural or artificial clustering of subjects (e.g., multiple teeth in subjects or individual participant data from multiple studies). For example, when trying to construct a prediction model using multiple data sources, it is inappropriate to ignore the dependence among study participants from the same study and simply analyze individual participant data from multiple studies as if they all came from a single

study(5). Finally, the standard Cox model treats all covariates that are included in the analysis as time-fixed, meaning that the values of covariates that change over time, such as kidney function, are not updated. While this is not an issue for covariates that are naturally time fixed, such as gender, keeping the values of time-varying covariates fixed at their baseline values has shown to result in an underestimation of the effect of that covariate on the time to occurrence of the event of interest, such as the onset of cardiovascular disease (CVD).

Advanced statistical approaches that aim to improve the predictive accuracy of clinical prediction models for time-to-event outcomes are well established in the medical statistical literature. However, those approaches are less frequently or incorrectly applied in the clinical literature. This thesis contributes toward bridging the gap between statistical and clinical research by using a series of clinical case studies to illustrate how sophisticated statistical models can be appropriately applied to obtain better predictions. Furthermore, this thesis contributes towards the medical statistical literature by empirically comparing the predictive performance of different dynamic risk prediction approaches.

Etiology versus prediction studies

In medical research, it is crucial to distinguish prediction studies from etiological studies since they involve different research questions. In etiological studies, the aim is to understand a certain pathway of disease, while in prediction studies the aim is

to predict a disease or disease prognosis as accurately as possible(6). As they do share some common methodology, they are frequently confused and poorly reported(7–10). For example, Lim et al. aimed to determine the impact of vascular disease burden on long-term transplantation and patient survival after kidney transplantation(11). This should have been an etiological study. However, the confounders selected in this study were based on p-values (predictive ability) in univariable analyses, which is an approach to select predictors in a prediction study. On the contrary, in another prediction study(12) aimed at developing a risk score to predict 5-, 10-, and 20-year individual dementia risk in older individuals, the authors gave a causal interpretation of the predictors included in the risk score by stating that "this risk estimate system helps individuals to identify their potential risk profile, and prevent or delay the future incidence of dementia".

In etiological studies, causality is usually of main interest. The gold standard for estimating causal effects is the randomized controlled trial (RCT) since a well-conducted RCT allows for a causal interpretation of the estimated treatment effect. In observational studies, causal relations are also estimable provided that certain non-observed confounding assumptions are being met(13,14). In prediction studies, the goal is to estimate individual's absolute risk of experiencing an event of interest using a combination of predictors(15). Candidate predictors included in these multivariable regression models are potentially associated with, but not necessarily causally related to, the outcome. Estimates of the strength of association between a predictor and the

outcome, such as an odds ratio, risk ratio, or hazard ratio, therefore have no direct implication in prediction studies. Instead, model accuracy is assessed in terms of measures related to discrimination (i.e., the ability to distinguish between subjects who have the outcome of interest and those who do not) and calibration (i.e., how well the predicted probability from the model agree with the observed outcome frequencies in the data) and these measures are again of little interest in etiological studies.

For a more detailed comparison between etiological and prediction research, we refer to Ramspek and his colleagues' scoping review(10). We primarily focus on prediction research in this thesis.

Static versus dynamic prediction

In this thesis, clinical prediction models are divided into two categories: static prediction models and dynamic prediction models. Static prediction models involve using a set of time-fixed predictors to estimate event risk (often termed overall survival) within a specific time period(16). Predictors' values at or before baseline are used and assumed to be constant across the follow-up. While prognostications are generally accurate at baseline, they tend to lose value for patients who have already survived for some time. In such situations, more accurate predictions can be optioned by updating the prognosis in response to changes in the clinical status of those patients(17,18). This is the domain of dynamic prediction.

One key feature differentiating dynamic prediction from static prediction is that the former usually involves time-dependent variables. There are two different types of time-dependent variables: internal time-dependent variables and external time-dependent variables(19). An external variable is one that is not directly related to the event process. For example, the level of air pollution is an external time-dependent variable when studying its association with asthma attacks. An internal variable is a value over time generated by the individual under study. Examples would include all the biomarkers such as blood pressures measured over the course of the study. The distinction between two types of time-dependent variable is helpful in choosing correct models in dynamic prediction. For example, the commonly used time-dependent Cox model (TDCM)(20) cannot properly handle internal variables because it assumes that the future measurements are independent of subjects' survival. In this thesis, we focus on dynamic prediction based on internal time-dependent variables.

Contemporary issues in static prediction

Researchers may encounter various challenges when developing and evaluating static prediction models(21). In this thesis, we focus on two common issues: between-study heterogeneity and competing risks.

Between-study heterogeneity inevitably exists when studies are brought together in a systematic review. When constructing a clinical prediction model by

synthesizing the evidence from multiple data sources, differences in study design, case-mix (i.e., different distributions of patient characteristics across studies), and follow-up period are all possible sources of variability that need to be accounted for in the estimation of the baseline risk and/or the specification of the predictor effects(22)(23). Previous methodological research has described and proposed several approaches for assessing and addressing heterogeneity in prediction models(22). These approaches have been applied in various fields including, but not limited to superficial bladder cancer(24), traumatic brain injury(25) and pulmonary embolism(26). However, they have not yet been applied in heart failure (HF), which is a heterogeneous syndrome from both an etiological and pathophysiologic standpoint(27). **Chapter 2** describes the development and validation of a prognostic model for predicting one-year all-cause mortality in patients hospitalized because of acute heart failure (HF) though an individual participant data (IPD) meta-analysis of four European acute HF cohorts.

Competing risks refer to one or more events whose occurrence preclude the event of interest (primary event) from happening. For example, when incident HF is the primary event, death acts as a competing risk as it precludes the new onset of HF from occurring. The issue of competing risks in clinical studies is prevalent. Austin and Fine reviewed how competing risks were addressed in RCTs published in four leading medical journals(28) and the results were rather alarming: 77.5% of RCTs with a survival outcome were potentially susceptible to competing risks, but only

16.1% of those studies properly addressed the competing risks. A similar conclusion was drawn by Koller et al. in their critical appraisal of 50 studies published in high-impact medical journals(29).

Since competing risks are typically informative, i.e., it will change the probability of event of interest, the key feature of the analysis of a time-to-event outcome in the presence of a competing risk is that the one-to-one relationship between hazard function and survival function (also referred to as the relationship between rate and risk in some studies(30,31)) is lost. In such situations, the standard Cox model, by treating competing events as censored observations, results in biased estimates of the absolute event risks(32,33). Although there is growing awareness of the impact of competing risks when developing prediction models, especially in cardiology(34–36) and nephrology(37–39), techniques for addressing competing risks (modeling of the cause-specific hazard functions and modeling of the subdistribution hazard function) are still not well understood and results are often poorly reported(33). **Chapter 3** describes the use of the Fine-Gray model to address the competing risk issue in the development of a prognostic model for predicting in-hospital mortality in COVID-19 patients.

Contemporary issues in dynamic prediction

As we introduced earlier, the defining feature of dynamic prediction is that predictions are updated in response to changes in the disease status of the patient.

Compared to assuming that all predictors remain constant across the follow-up as in static prediction, dynamic prediction is statistically more demanding as it requires modeling of the change in the (distribution of the) predictor variable over the time horizon of the prediction model. That is to say, dynamic prediction requires knowledge of the predictor's values at future time points. For example, if one is interested in using dynamic prediction to estimate a patient's 5-year probability of developing CVD based on time-varying systolic blood pressure (SBP) values, these SBP values need to be known 5 years forward.

One way to obtain the future value of a predictor is to explicitly model the trajectory of that predictor. This is the approach taken in the shared random effects model (SREM)(40), which is a standard model for the joint modelling of longitudinal and survival data. In the SREM, a linear mixed effects model is used to model the trajectory of the predictor variable while the association between the present value of that predictor and the risk of experiencing the event of interest is modeled using a parametric or semi-parametric (Cox-like) survival model. While the SREM is extensively researched in the medical statistics field, it is still less frequently applied in the clinical field. **Chapter 4** introduces the SREM and stresses its advantages over the commonly used TDCM through an illustrative case study in respiratory medicine.

An alternative approach to jointly model longitudinal and survival data is the hidden Markov model (HMM)(41). The HMM assumes that a patient's prognosis depends on the underlying state of the disease, which cannot be directly observed

(i.e., is latent). The observed longitudinal data are assumed to be realizations from a set of probability distributions conditional on these latent states. As such, the likelihood of a patient residing in a particular state can be inferred from the available longitudinal data, making the HMM a suitable tool for dynamic prediction. The HMM is commonly used in areas such as speech and signal processing(42) and has also been successfully applied to biological sequences(43). However, the HMM is not widely applied to clinical studies. **Chapter 5** comprehensively assesses the association between CKD stage and the development of HF by fitting a misclassification model, an important special case of the HMM(41), to the data from the PREVEND study.

It is natural to think that dynamic prediction models, by including the change of some important predictors, can achieve more accurate predictions compared with those including baseline predictors only. This has been confirmed in several clinical fields including but not limited to diabetes(44), breast cancer(45), and cystic fibrosis(46). However, the comparison of the predictive performance of different dynamic prediction approaches is less researched. While previous research showed that the SREM outperformed more naive approaches such as landmarking(47,48), it is unclear how the HMM, a less frequently applied approach in dynamic prediction, performs compared to the SREM. In **Chapter 6**, the predictive performance of the SREM and the HMM are empirically evaluated in the context of dynamically

predicting mortality in patients with acute HF based on serial NT-proBNP measurements.

References

1. Baart SJ, van den Berge JC, Akkerhuis KM, Deckers JW, van Domburg RT, Boersma E, et al. Relative conditional survival analysis provides additional insights into the prognosis of heart failure patients. European Journal of Preventive Cardiology [Internet]. 2021 Feb 1 [cited 2021 Aug 10];(zwab003). Available from: https://doi.org/10.1093/eurjpc/zwab003

2. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457–81.

3. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology. 2009 Jul;20(4):555–61.

4. Gharibv, LidaLiu L. 237-2009: Analysis of Survival Data with Clustered Events. 2009;11.

5. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010 Feb 5;340:c221.

6. van Diepen M, Ramspek CL, Jager KJ, Zoccali C, Dekker FW. Prediction versus aetiology: common pitfalls and how to avoid them. Nephrology Dialysis Transplantation. 2017 Apr 1;32(suppl_2):ii1–5.

7. Hayden JA, Côté P, Steenstra IA, Bombardier C, QUIPS-LBP Working Group. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. J Clin Epidemiol. 2008 Jun;61(6):552–60.

8. Herbert RD. Cohort studies of aetiology and prognosis: they're different. J Physiother. 2014 Dec;60(4):241–4.

9. Shmueli G. To Explain or to Predict? Statistical Science. 2010 Aug;25(3):289–310.

10. Ramspek CL, Steyerberg EW, Riley RD, Rosendaal FR, Dekkers OM, Dekker FW, et al. Prediction or causality? A scoping review of their conflation within current observational research. Eur J Epidemiol. 2021;36(9):889–98.

11. Lim WH, Johnson DW, Hawley CM, Pascoe E, Wong G. Impact of Diabetes Mellitus on the Association of Vascular Disease Before Transplantation With

Long-term Transplant and Patient Outcomes After Kidney Transplantation: A Population Cohort Study. Am J Kidney Dis. 2018 Jan;71(1):102–11.

12. Li J, Ogrodnik M, Devine S, Auerbach S, Wolf PA, Au R. Practical risk score for 5-, 10-, and 20-year prediction of dementia in elderly persons: Framingham Heart Study. Alzheimers Dement. 2018 Jan;14(1):35–42.

13. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. Journal of Epidemiology & Community Health. 2006 Jul 1;60(7):578–86.

14. Instruments for Causal Inference: An Epidemiologist's Dream? : Epidemiology [Internet]. [cited 2022 May 16]. Available from: https://journals.lww.com/epidem/Fulltext/2006/07000/Instruments_for_Causal_Inference__An.4.aspx

15. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ [Internet]. 2009 Feb 23 [cited 2020 May 4];338. Available from: https://www.bmj.com/content/338/bmj.b375

16. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012 May 1;98(9):683–90.

17. Schumacher M, Hieke S, Ihorst G, Engelhardt M. Dynamic prediction: A challenge for biostatisticians, but greatly needed by patients, physicians and the public. Biom J. 2020 May;62(3):822–35.

18. Houwelingen H van, Putter H, Putter H. Dynamic Prediction in Clinical Survival Analysis [Internet]. CRC Press; 2011 [cited 2019 Oct 15]. Available from: https://www.taylorfrancis.com/books/9780429094330

19. Likelihood Construction and Further Results. In: The Statistical Analysis of Failure Time Data [Internet]. John Wiley & Sons, Ltd; 2002 [cited 2021 Sep 18]. p. 193–217. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118032985.ch6

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

21. White IR, Wood AM. Statistical Challenges in Risk Prediction. Biom J. 2015 Jul;57(4):529–30.

22. Steyerberg EW, Nieboer D, Debray TPA, van Houwelingen HC. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An overview and illustration. Statistics in Medicine. 2019;38(22):4290–309.

23. Abo-Zaid G, Guo B, Deeks JJ, Debray TPA, Steyerberg EW, Moons KGM, et al. Individual participant data meta-analyses should not ignore clustering. J Clin Epidemiol. 2013 Aug;66(8):865-873.e4.

24. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. Statist Med. 2004 Mar 30;23(6):907–26.

25. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008 Aug 5;5(8):e165; discussion e165.

26. Geersing GJ, Takada T, Klok FA, Büller HR, Courtney DM, Freund Y, et al. Ruling out pulmonary embolism across different healthcare settings: A systematic review and individual patient data meta-analysis. PLOS Medicine. 2022 Jan 25;19(1):e1003905.

27. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. European Journal of Heart Failure. 2020;22(8):1342–56.

28. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. Stat Med. 2017 Apr 15;36(8):1203–9.

29. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? Stat Med. 2012 May 20;31(11–12):1089–97.

30. Competing Risk Regression Models for Epidemiologic Data | American Journal of Epidemiology | Oxford Academic [Internet]. [cited 2021 Nov 26]. Available from: https://academic.oup.com/aje/article/170/2/244/111339

31. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012 Jun;41(3):861–70.

32. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999 Jun;94(446):496–509.

33. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017 Nov 30;36(27):4391–400.

34. Introduction to the Analysis of Survival Data in the Presence of Competing Risks | Circulation [Internet]. [cited 2021 Nov 26]. Available from: https://www.ahajournals.org/doi/10.1161/circulationaha.115.017719

35. Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondré K, et al. Competing risks analyses: objectives and approaches. Eur Heart J. 2014 Nov 7;35(42):2936–41.

36. Postmus D, Demissei BG, Hillege HL. Tools and Techniques - Statistical. Estimating cumulative incidences in the presence of right-censoring and competing risks: an introduction with illustrations from the COACH study. EuroIntervention. 2016 Feb;11(11):e1322-1324.

37. Al-Wahsh H, Tangri N, Quinn R, Liu P, Ferguson M Thomas, Fiocco M, et al. Accounting for the Competing Risk of Death to Predict Kidney Failure in Adults With Stage 4 Chronic Kidney Disease. JAMA Network Open. 2021 May 4;4(5):e219225.

38. Li L, Yang W, Astor BC, Greene T. Competing Risk Modeling: Time to Put it in Our Standard Analytical Toolbox. J Am Soc Nephrol. 2019 Dec;30(12):2284–6.

39. Noordzij M, Leffondré K, van Stralen KJ, Zoccali C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? Nephrology Dialysis Transplantation. 2013 Nov 1;28(11):2670–7.

40. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data : With Applications in R [Internet]. Chapman and Hall/CRC; 2012 [cited 2019 Oct 7]. Available from: https://www.taylorfrancis.com/books/9780429063381

41. Jackson CH. Multi-State Models for Panel Data: The **msm** Package for *R*. J Stat Soft [Internet]. 2011 [cited 2021 Feb 8];38(8). Available from: http://www.jstatsoft.org/v38/i08/

42. Juang BH, Rabiner LR. Hidden Markov Models for Speech Recognition. Technometrics. 1991;33(3):251–72.

43. Yoon BJ. Hidden Markov Models and their Applications in Biological Sequence Analysis. Curr Genomics. 2009 Sep;10(6):402–15.

44. Parast L, Mathews M, Friedberg MW. Dynamic risk prediction for diabetes using biomarker change measurements. BMC Med Res Methodol. 2019 Dec;19(1):175.

45. 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.

46. 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.

47. Rizopoulos D, Molenberghs G, Lesaffre EMEH. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biometrical Journal. 2017 Nov;59(6):1261–76.

48. Ferrer L, Putter H, Proust-Lima C. Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. Stat Methods Med Res. 2019 Dec;28(12):3649–66.

Part I: Contemporary issues in static prediction