Contemporary issues in static and dynamic prediction

Chen, Yuntao

DOI:
10.33612/diss.232289260

Chapter 6
Dynamic risk prediction for all-cause mortality in patients with acute heart failure: an empirical evaluation of the shared random effects model against the hidden Markov model

Yuntao Chen, Gerton Lunter, Eric Boersma, Hans L. Hillege, Douwe Postmus

submitted

Abstract

Background Dynamic prediction by incorporating repeated biomarker measurements over time might improve predictive accuracy. This study aimed to perform an empirical evaluation of two approaches for dynamically predicting all-cause mortality based on serial NT-proBNP measurements in patients with acute heart failure (HF).

Methods The TRanslational Initiative on Unique and novel strategies for Management of Patients with Heart failure (TRIUMPH) cohort was analyzed with all-cause mortality as the outcome of interest. Two dynamic prediction approaches, namely joint modelling of longitudinal and survival data based on a shared random effects model (SREM) and the hidden Markov model (HMM), were applied to the TRIUMPH cohort and their predictive performance was compared in terms of the area under the cumulative/dynamic time-dependent ROC curve (AUC) and the Brier Score (BS).

Results 421 patients were included in our analysis, of whom 99 (23.5%) reached the all-cause mortality endpoint within the 400 day follow-up period of the TRIUMPH cohort. The performance of the SREM and the HMM was similar in the unadjusted analysis. In the adjusted analyses, the AUCs were higher for the HMM than for the SREM. In terms of BS, the SREM performed better at a window time of 90 days while the HMM performed better at a window time of 180 days.

Conclusions The HMM performed well relative to the SREM and could therefore become a useful tool to aid dynamic prediction in clinical research.

Keywords Dynamic prediction; Hidden Markov model; NT-proBNP; Heart failure

Introduction

Heart failure (HF) is a heterogeneous syndrome with an increasing global burden(1). Accurately predicting prognosis could help patients and physicians to decide on the type and timing of therapy and assist with the planning of social and medical resources(2).

Many prediction models are available to estimate an acute HF patient's prognosis at hospital admission or discharge. Throughout the clinical follow-up, these predictions will have to be updated to reflect changes in the patient's disease status or to take into account the fact that the patient has survived up to a specific time point (e.g., after hosptitalization because of acute HF, mortality is much higher in the first year than in the subsequent follow-up(3)). A straightforward way to obtain dynamic predictions is to calculate conditional survival probabilities from a Cox proportional hazards model evaluated with the information at baseline. A downside of this approach is that it does not consider that a patient's disease status at the time of updating may be different from that patient's disease status at baseline, possibly resulting in poor predictions at later stages in the follow-up. In such situations, joint modelling of the available longitudinal and survival data may result in more accurate dynamic risk prediction(4).

The shared random effects model (SREM)(5) is a standard model for the joint modelling of longitudinal and survival data. This approach, which is based on the assumption that a subject's longitudinal and survival data are conditionally

independent given a vector of shared random effects, has been shown to result in better predictive performance compared to more naive approaches, such as conditional survival probabilities derived from a baseline Cox model or landmarking(6,7). An alternative approach to jointly model the available longitudinal and survival data is the hidden Markov model (HMM)(8). 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 the latent state. 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. Information on the predictive performance of the HMM compared to that of the SREM is limited, however. The objective of this study was to empirically evaluate the performance of the HMM against that of the SREM in the context of dynamically predicting all-cause mortality for patients with acute HF based on serial NT-proBNP measurements.

Background

Estimand

The purpose of dynamic prediction is to estimate the probability that a subject survives an additional w units of time starting from a time point t_{LM} at which they are still event free, i.e.,

$$\pi_i(t|t_{LM}) = P(T_i > t|T_i > t_{LM}, Z_i, y_i(t_{LM})),$$

where $t = t_{LM} + w$ is the horizon time, Z_i a vector of time-fixed predictors at baseline, and $y_i(t_{LM})$ the history of longitudinal measurements up to time t_{LM}. The time point t_{LM} is referred to as the landmark time and the time period w as the window time.

Baseline Cox model

In situations where all predictors are being treated as time-fixed, dynamic predictions can readily be obtained by using the following formula for calculating conditional survival (CS) probabilities(9):

$$\hat{\pi}(t|t_{LM}) = \hat{S}(t)/\hat{S}(t_{LM}),$$

where $\hat{S}(t)$ is an estimated survival function obtained using standard approaches such as the Kaplan-Meier estimator (no predictors) or Cox proportional hazards regression (baseline information only). In this study, the baseline Cox model is used to set a benchmark for the other more advanced models.

SREM

The SREM is essentially linking a longitudinal process to a survival process, which includes two sub-models: a longitudinal sub-model and a survival sub-model(10). For the longitudinal sub-model, a linear mixed model is typically used to describe the subject-level longitudinal trajectory, which is given as follows:

$$y_i(t) = m_i(t) + \varepsilon_i(t),$$

$$m_i(t) = a_i + b_i t,$$

where a_i and b_i are the intercept and slope (modelled as random coefficients that are distributed according to a bivariate normal distribution) of the biomarker trajectory of the i-th subject. $\varepsilon_i(t)$ is a normally distributed error term representing the random fluctuation of the observed measurements $y_i(t)$ around the modeled marker trajectory $m_i(t)$. For the survival sub-model, a relative risk model (e.g. Cox model) is used to quantify the association between the biomarker trajectory and the risk for an event of interest, which is given as follows:

$$h_i(t) = h_0(t)\exp(\alpha Z_i + \beta m_i(t)),$$

where $m_i(t)$ is the modelled longitudinal process up to t; $h_0(t)$ is the baseline hazard function; Z_i is a vector of baseline covariates with corresponding regression coeffcicient α; and β quantifies the association between the true value of the marker at time t and the hazard for the event at the same time.

The different components of the SREM (i.e., the parameters of the hazard function, the parameters of the bivariate normal distribution of the intercepts and slopes of the subject-level biomarker trajectories, and the standard error of the normally distributed error term) are jointly estimated from the available longitudinal and survival data by maximizing the joint likelihood of these observations. For a more detailed discussion of the construction of the likelihood function and the available estimation procedure, we refer to Rizopoulos(10).

After fitting the model, the first-order estimate of $\pi_i(t|t_{LM})$ can be obtained using the empirical Bayes estimator, i.e.,

$\hat{\pi}_i(t|t_{LM}) = \Pr\,(t|t_{LM}, \hat{a}_i, \hat{b}_i, \hat{\theta}, Z_i)$,

where $\hat{a}_i$ and $\hat{b}_i$ denotes the empirical Bayes estimates; $\hat{\theta}$ denotes the maximum likelihood estimates of the model parameters.

HMM

HMMs can be seen as an alternative for the SREM. It is a direct extension of the Markov chain with a stochastic observation process. The underlying Markov chain is not observable while the observed longitudinal data are governed by some probabilistic distribution conditional on the unobserved state(8). It can also be considered as having two sub-models: a survival sub-model and a longitudinal sub-model. For the survival sub-model, a multi-state model is used to describe the stochastic Markov process for latent states and absorbing state. For a continuous-time Markov chain $S(t)$ on finite state space N, time-homogeneous transition probabilities are given by

$p_{gh}(t) = P(S(t + \Delta t) = h|S(t) = g)$,

where $g, h \in N$. The probability of being in state h at time $t + \Delta t$ given the current state g at time t depends only on the elapsed time t. Transition probability matrix $\boldsymbol{P}(t)$ contains these probabilities with the rows sum equalling to 1. The Chapman-Kolmogorov equation is $\boldsymbol{P}(t + \Delta t) = \boldsymbol{P}(t)\boldsymbol{P}(\Delta t)$. The transition intensities

representing the instantaneous risk of moving from state g to state $h \neq g$ is as follows:

$$q_{gh}(t) = \lim_{\Delta t \to 0} P(S(t + \Delta t) = h | S(t) = g)/\Delta t,$$

The transition intensity can also depend on other covariates, given by:

$$q_{gh}\big(t, z(t)\big) = q_{gh}^{(0)} \exp\left(\beta_{gh} z_{gh}\right),$$

where $q_{gh}^{(0)}$ is the baseline intensity for transition g to h; z_{gh} is transition-specific covariates with corresponding regression coefficients β_{gh}. For the longitudinal sub-model, the distribution of the longitudinal measurements is conditional on the state defined above, coming from a parametric family of continuous or discrete distributions. Therefore, the unknown parameters in an HMM involve both the parameters of the Markov chain and those of the state-dependent distributions of the longitudinal observations. Estimation of model parameters is undertaken by maximising the log-likelihood function. For a more detailed discussion of the construction of the likelihood function and the available estimation procedure, we refer to Jackson(8).

After fitting the HMM, the estimate of $\pi_i(t|t_{LM})$ can be calculated using the forward algorithm(11) given the subject's longitudinal measurements up to t_{LM} and the information that subject is still alive at t_{LM}. The 'forward' function in 'HMM' package was referred to obtain the estimate(11).

Methods

Data source

The present study was performed using data from the TRanslational Initiative on Unique and novel strategies for Management of Patients with Heart failure (TRIUMPH) cohort, a prospective cohort study that enrolled patients admitted with acute HF in 14 hospitals in the Netherlands between September 2009 and December 2013. TRIUMPH was designed to identify and validate potentially clinically important biomarkers that could be implemented in the clinical management of acute HF patients to monitor disease progression and therapeutic efficacy, as well as to sub-diagnose and guide tailored therapy. Details of the study protocol have been described elsewhere. The study was approved by the medical ethics committees at all participating centers.

Outcome and predictors

In our study, the outcome of interest was all-cause mortality. The follow-up was defined as the time from the date of hospital admission to the date of death from any cause, or 400 days after hospital admission, whichever came first. For simplicity, we only focused on one time-dependent predictor, NT-proBNP, which remains the gold standard biomarker in HF(12). NT-proBNP was measured at admission (day 1), once during days 2-4, and subsequently on the day of discharge. Afterwards, measurements were also planned at 2 to 4 weeks, 3 months, 6 months and 9 to 12

months after discharge. NT-proBNP concentrations were determined in heparin plasma by using the Elecsys NT-proBNP electro-chemiluminescent sandwich immunoassay on a Cobas 8000 analyzer (Roche Diagnostics, Ltd., Rotkreuz, Switzerland). Other variables measured at hospital admission that were used as adjusters in our analysis included age, sex, myocardial infarction (MI), COPD, diabetes, New York Heart Association (NYHA) class, systolic blood pressure (SBP), haemoglobin, serum sodium, serum creatinine, and blood urea nitrogen (BUN).

Study design

We aimed to derive subject-specific conditional survival predictions at the landmark times of 30, 60, 90, 120, 150, and 180 days. The window times were set to 90 and 180 days. For example, with a window time of 90 days and a landmark time of 30 days, the goal is to predict the probability that a patient is still alive at day 120 given that they were still alive at day 30. For each of the three dynamic prediction approaches, three subsequently more complex models were estimated: 1) an unadjusted model that only included the serial NT-proBNP measurements; 2) a clinical model(13) including NT-proBNP, age, SBP, serum sodium, NYHA class, and BUN; 3) a full model(14) additionally including MI, COPD, diabetes, haemoglobin, and serum creatinine. Missing values for the predictor variables were handled using single imputation with the ‘mice’ package in R(15).

Estimation of the baseline Cox model

For the baseline Cox model, we fit a Cox proportional hazards models to the measurements at baseline according to three risk prediction equations. We then obtained individual predictions using the CS function. For example, in the setting of the clinical model, we first fitted a Cox model with NT-proBNP, age, SBP, serum sodium, NYHA class at baseline. To obtain a prediction for a subject's 120 days survival probability given that they are alive at day 30 day, we calculated their 30 day and 120 day survival probability based on the fitted Cox model. The survival probability of interest can then be obtained by dividing the 120 day survival probability by the 30 day survival probability.

Estimation of the SREMs

We fit a SREM by combining a linear mixed model for serial NTpro-BNP measurements and a Cox proportional hazards model for the mortality risk. Specifically, we fit a mixed model with random intercept and random slope. Follow-up time was the only variable included in the mixed model. We applied cubic spline function for follow-up time, with knots set at one week and one month after hospital admission as what has been done previously(16). For the Cox model, we again fit three models according to three risk prediction equations. After fitting the model, we obtained conditional survival probabilities according to different settings using the 'survfitJM' function in R package 'JM'(17).

Estimation of the HMMs

We represented the HF progression by a HMM for NT-proBNP, conditional on underlying HF states. For the survival model, we considered a multi-state model with three underlying HF states and death (Figure 1). Initially we considered a time-homogenous exponential hazard model with the following transition intensity matrix:

$$Q = \begin{pmatrix} -(q_{12}+q_{13}+q_{14}) & q_{12} & q_{13} & q_{14} \\ q_{21} & -(q_{21}+q_{23}+q_{24}) & q_{23} & q_{24} \\ 0 & q_{32} & -(q_{32}+q_{34}) & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

The effect of covariates on the transition intensities were modelled according to three risk prediction equations. For simplicity, covariates were only included on the transition to death. We also assumed proportional assumptions, i.e. the effects of covariate on the three transition intensities q_{14}, q_{24} and q_{34} were the same. Conditional on the three hidden HF states, we assumed NT-proBNP measurements followed log-normal distributions as follows:

$\log(Y_1) \sim N(\mu_1, {\sigma_1}^2)$,

$\log(Y_2) \sim N(\mu_2, {\sigma_2}^2)$,

$\log(Y_3) \sim N(\mu_3, {\sigma_3}^2)$.

Death was considered to be observed without error. A likelihood ratio test was conducted to determine whether constructing a time-inhomogeneous model with piecewise-constant intensities changing at 60 and 180 days could improve the model fit. After fitting the HMM, we obtained predictions according to different settings.

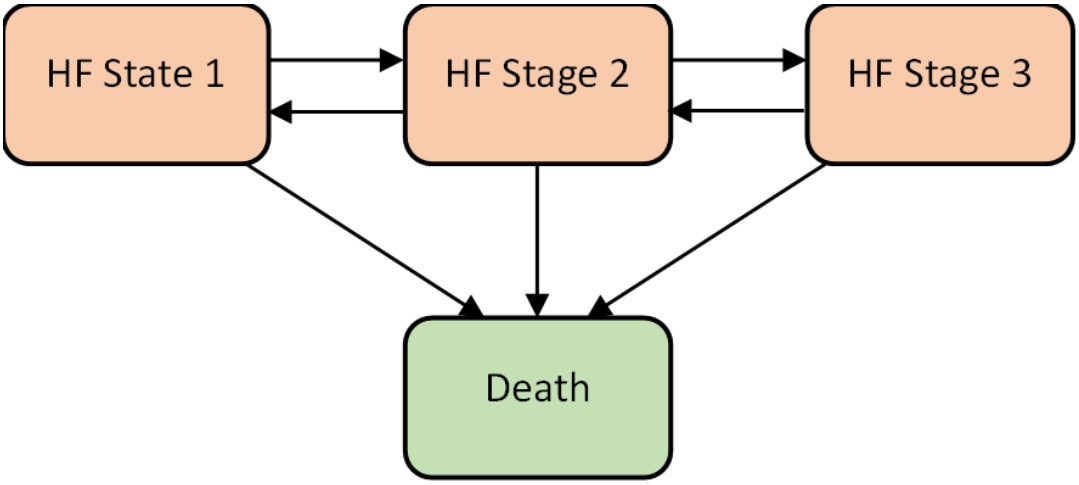

Figure 1. The Hidden Markov model with three heart failure hidden states and one absorbing state death.

Model assessment

Models were assessed in terms of discrimination and calibration using a five-fold cross-validation approach. Discrimination was assessed using the area under the cumulative/dynamic time-dependent ROC (AUC) computed at the evaluated landmarking time point, using the nonparametric inverse probability of censoring weighted version of the estimator proposed by Blanche et al.(18). Calibration was assessed using the Brier Score (BS)(19), which is defined as:

$BS(t, t_{LM}) = E([N_i(t) - \hat{\pi}_i(t|t_{LM})]\text{^}2)$,

where $N_i(t) = I(T_i \leq t)$ is the event status at time t; $\hat{\pi}_i(t|t_{LM})$ is the predicted event probability at time t. The BS captures the mean squared error comparing the

true event rates and the predicted event rates obtained from the prediction model and was estimated using the nonparametric inverse probability of censoring weighted approach suggested by Gerds and Schumacher(20). Because the BS depends on the marginal survival distribution, we applied the following rescaling:

$BS_{rel} = (BS_{KM} - BS_{model})/BS_{KM}$,

where BS_{KM} is the BS of the Kaplan-Meier estimate of the event probability (21,22). Higher values of BS_{rel} indicate better calibration, with $BS_{rel} > 0$ indicating smaller prediction errors than that of a null model involving the Kaplan-Meier estimator.

Results

At baseline, 421 patients were included in our sample, with a median age of 73 and 36.6% of female. 69 (16.4%) patients were in NYHA class I/II, 231 (54.9%) were in NYHA class III, and 102 (24.2%) were in NYHA class IV. 148 (35.2%) patients had diabetes, 80 (19%) had COPD, and 171 (40.6%) had myocardial infarction (Table 1). 99 (23.5%) patients died within the follow-up of 400 days, with most of them died in the first 120 days. Most of the patients experienced a drop for NT-proBNP level at the beginning of the follow-up, with the dead patients experiencing an averaged increasing trend while the alive patients flattening afterwards (Figure 2).

Table 1. Baseline characteristics

	Overall sample (n=421)
Age, years	73 (64,80)
Sex, female	154 (36.6%)
NYHA class	
I/II	69 (16.4%)
III	231 (54.9%)
IV	102 (24.2%)
SBP, mmHg	125 (110,147)
Diabetes	148 (35.2%)
COPD	80 (19.0%)
Myocardial infarction	171 (40.6%)
Hemoglobin,	8.10 (7.20,9.10)
Serum sodium	139 (137,141)
Serum creatinine	114 (87,152)
Blood urea nitrogen	9.65 (6.98,14.1)
NT-proBNP, pg/ml	4316 (2165,9509)

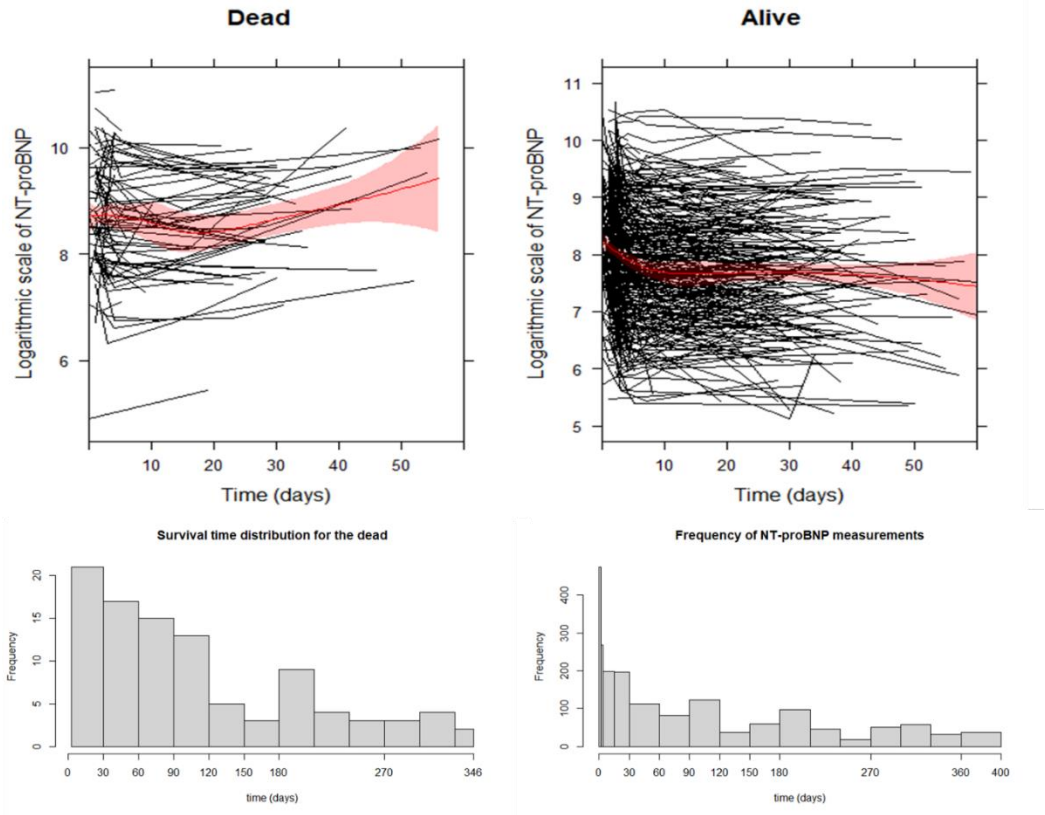

Figure 2. longitudinal trajectory of log-transformed NT-proBNP for dead and alive patients; Distribution of survival time for the dead patients; Frequency of NT-proBNP measurements across follow-up time.

Comparison of predictive performances

Window time 90 days

In the unadjusted model setting, the AUCs of HMM and SREM both performed well and were much better than the baseline Cox model. In the clinical model setting, the AUC of HMM was the best, especially at the landmark time of 120 and 150 days. The AUCs of HMM and SREM were both better than that of the baseline Cox model, with a smaller difference than that in the univariable model setting. In the full model

setting, we observed the same trend of AUCs as in the clinical model setting, with the HMM having the overall best performance (Figure 3; Table 2).

In all the three model settings, the BSs were not much better than the null model for all the approaches at earlier landmark time points (30, 60 and 90 days). At later landmark time points (120, 150 and 180 days), the relative error reduction of HMM and SREM were better than that of baseline Cox model. These differences were larger in clinical and full model settings than in univariable model setting. The relative error reduction of SREM was slightly better than that of HMM (Figure 3).

Window time 180 days

At the window time of 180 days, we observed the same trend of AUCs as that observed at the window time of 90 days, with the HMM having the overall best performance (Figure 4; Table 3).

However, in the window time of 180 days, we observed a different trend of relative error reduction, with increased relative error reductions along with the larger landmark time points for all the approaches. Though the relative error reductions of HMM and SREM were better than that of baseline Cox model, the differences were smaller than those in the window time of 90 days. Overall, HMM had the best relative error reduction (Figure 4).

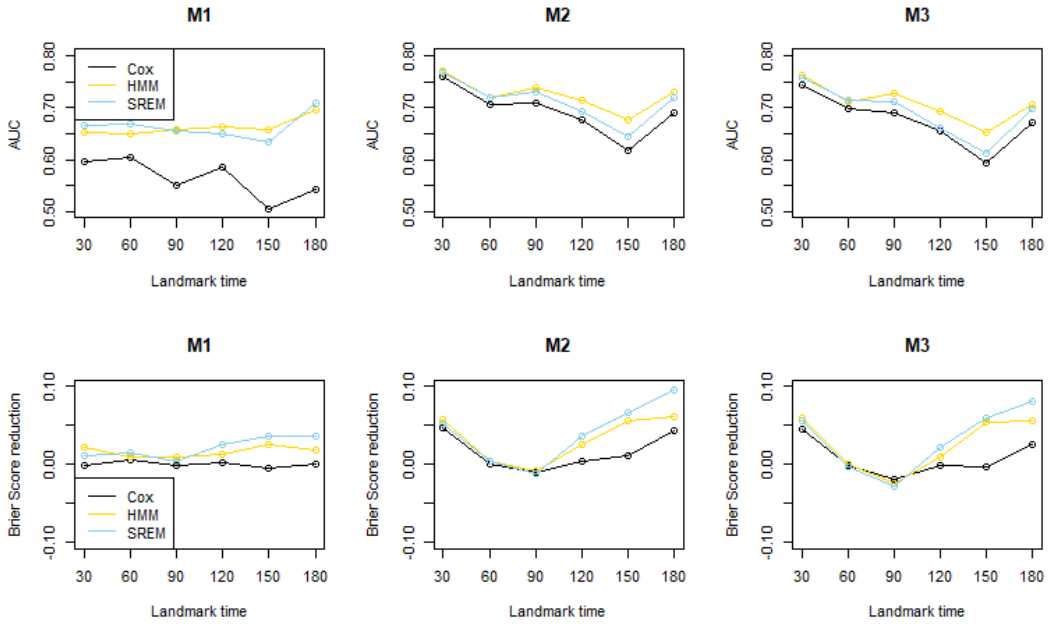

Figure 3. Estimated AUCs and relative Brier Scores for four dynamic prediction approaches baseline Cox model (Cox), Hidden Markov model (HMM) and shared random effect model (SREM) at three settings of model complexity (M1: univariable model; M2: clinical model; M3: full model) for window time of 90 days.

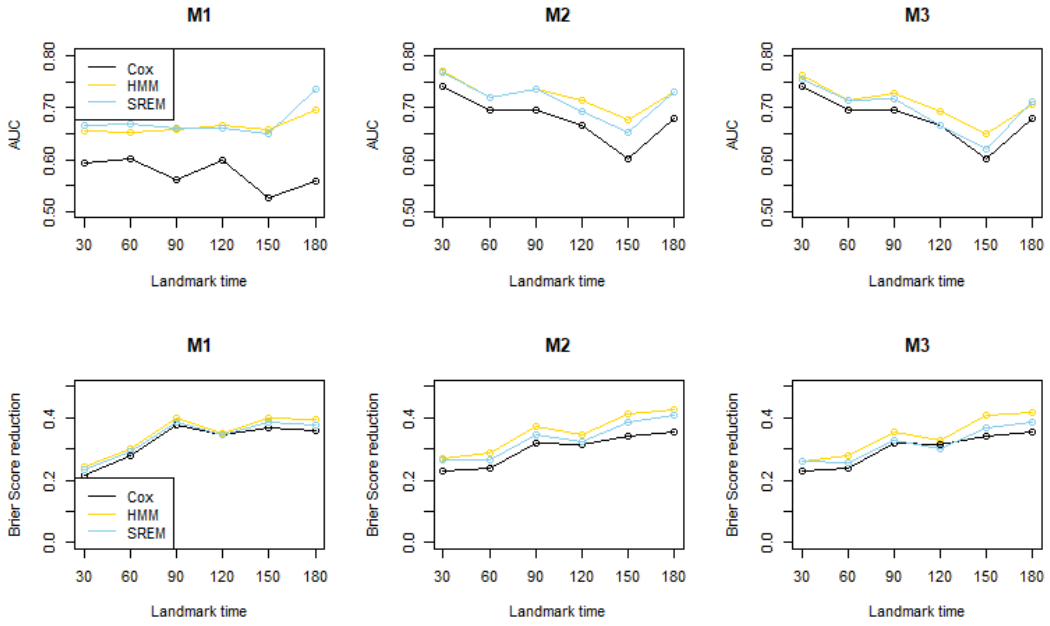

Figure 4. Estimated AUCs and relative Brier Scores for four dynamic prediction approaches baseline Cox model (Cox), Hidden Markov model (HMM) and shared random effect model (SREM) at three settings of model complexity (M1: univariable model; M2: clinical model; M3: full model) for window time of 180 days

Table 2. Comparison of AUCs for the HMM and SREM in different landmark time points (30, 60, 90, 120, 150, 180 days). The window time is 90 days.

	30			60			90			120			150			180		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Cox	0.595	0.759	0.744	0.605	0.707	0.697	0.551	0.709	0.691	0.585	0.676	0.656	0.505	0.618	0.594	0.543	0.691	0.672
SREM	0.667	0.768	0.755	0.667	0.719	0.713	0.656	0.729	0.711	0.651	0.691	0.661	0.633	0.646	0.611	0.706	0.719	0.699
HMM	0.654	0.770	0.762	0.651	0.719	0.713	0.658	0.738	0.728	0.664	0.714	0.694	0.657	0.677	0.653	0.709	0.730	0.706

Table 3. Comparison of AUCs for the HMM and SREM in different landmark time points (30, 60, 90, 120, 150, 180 days). The window time is 180 days.

	30			60			90			120			150			180		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Cox	0.595	0.758	0.742	0.603	0.704	0.696	0.561	0.711	0.695	0.598	0.683	0.666	0.527	0.635	0.603	0.558	0.703	0.679
SREM	0.666	0.768	0.755	0.668	0.720	0.714	0.661	0.734	0.717	0.659	0.694	0.665	0.649	0.653	0.619	0.737	0.730	0.712
HMM	0.654	0.771	0.762	0.651	0.719	0.713	0.659	0.736	0.727	0.666	0.713	0.694	0.658	0.677	0.651	0.697	0.731	0.706

Discussion

Using TRIUMPH cohort data, we empirically evaluated the predictive performance of the HMM against that of the SREM in the context of dynamically predicting all-cause mortality using 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 this patient cohort.

Previous studies have shown that serial NT-proBNP measurements had additional value in predicting the prognosis of HF(4,23). However, most of them were etiology studies investigating the association between serial NT-proBNP measurements and prognosis, and were not prediction studies investigating how prediction could be improved by inclusion of serial NT-proBNP measurements that has accumulated up to the start time of the prediction. Our study found that two dynamic prediction approaches (the HMM and SREM) incorporating serial NT-proBNP measurements performed better than the approach (the baseline Cox model) incorporating baseline NT-proBNP only, confirming the additional value of serial NT-proBNP in dynamic prediction in this patient cohort.

Our study introduced a less frequently used approach HMM to the field of dynamic prediction and compared its predictive performance with SREM. We found the predictive performance of HMM to be very robust and superior to SREM. This is promising, and suggests that the HMM could be a good alternative to SREM in

dynamic prediction. The main advantage of HMM over SREM lies in its intuitive interpretation. In chronic staged diseases such as HF, it is more reasonable to consider disease course as discrete states governed by biomarkers. Also, compared with the SREM, HMMs can be more easily extended to use measurements of more than one serial biomarker, which may further improve predictive performance(24). Limitations of the HMMs include the choice of the most appropriate number of states that is not always clear(25). Models with different number of states can disagree in their predictive performance.

Our study has several limitations. Firstly, the follow-up time was comparatively short. This may influence the predictive performance, especially for landmarking approach where fewer patients were at risk at later landmark time points. Secondly, our results were based on the TRIUMPH cohort which includes predominantly Caucasian patients with reduced ejection fraction, limiting the generalization to other populations.

Conclusion

Dynamic prediction by incorporating serial biomarker measurements could improve predictive performance. HMMs perform well compared with the SREM. It can be a useful tool to aid dynamic prediction in clinical research.

References

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

2. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal. 2016 Jul 14;37(27):2129–200.

3. 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

4. Zhang J, Pellicori P, Pan D, Dierckx R, Clark AL, Cleland JGF. Dynamic risk stratification using serial measurements of plasma concentrations of natriuretic peptides in patients with heart failure. International Journal of Cardiology. 2018 Oct;269:196–200.

5. Andrinopoulou ER, Harhay MO, Ratcliffe SJ, Rizopoulos D. Reflections on modern methods: Dynamic prediction using joint models of longitudinal and time-to-event data. Int J Epidemiol. 2021 Mar 17;dyab047.

6. 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.

7. 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.

8. 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/

9. Hieke S, Kleber M, König C, Engelhardt M, Schumacher M. Conditional Survival: A Useful Concept to Provide Information on How Prognosis Evolves over Time. Clin Cancer Res. 2015 Apr 1;21(7):1530–6.

10. 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

11. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989 Feb;77(2):257–86.

12. Sarhene M, Wang Y, Wei J, Huang Y, Li M, Li L, et al. Biomarkers in heart failure: the past, current and future. Heart Fail Rev. 2019 Nov;24(6):867–903.

13. Felker GM, Leimberger JD, Califf RM, Cuffe MS, Massie BM, Adams KF, et al. Risk stratification after hospitalization for decompensated heart failure. J Card Fail. 2004 Dec;10(6):460–6.

14. 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.

15. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec 12;45(1):1–67.

16. van Vark LC, Lesman-Leegte I, Baart SJ, Postmus D, Pinto YM, Orsel JG, et al. Prognostic Value of Serial ST2 Measurements in Patients With Acute Heart Failure. Journal of the American College of Cardiology. 2017 Nov;70(19):2378–88.

17. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software (Online). 2010 Jul 1;35(9):1–33.

18. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in Medicine. 2013;32(30):5381–97.

19. Brier GW. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. Monthly Weather Review. 1950 Jan 1;78(1):1–3.

20. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biometrical Journal. 2006;48(6):1029–40.

21. 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

22. Schumacher M, Binder H, Gerds T. Assessment of survival prediction models based on microarray data. Bioinformatics. 2007 Jul 15;23(14):1768–74.

23. Baggen VJM, Baart SJ, van den Bosch AE, Eindhoven JA, Witsenburg M, Cuypers JAAE, et al. Prognostic Value of Serial N-Terminal Pro-B-Type Natriuretic Peptide Measurements in Adults With Congenital Heart Disease. J Am Heart Assoc. 2018 Mar 26;7(7):e008349.

24. van den Hout A, Muniz-Terrera G. Hidden three-state survival model for bivariate longitudinal count data. Lifetime Data Anal. 2019 Jul;25(3):529–45.

25. Thom HHZ, Jackson CH, Commenges D, Sharples LD. State selection in Markov models for panel data with application to psoriatic arthritis: State selection in Markov models for panel data with application to psoriatic arthritis. Statist Med. 2015 Jul 20;34(16):2456–75.