Protocol for Comparative Methods Study on Penalisation Methods for Multi-State Models

Ms. Chantelle Cornett (CC), Dr. Glen Martin (GM), Dr. Niels Peek (NP)

Overview

Multi-state models (MSMs) are useful for prediction of complex clinical settings where patients may transition through several intermediate states before experiencing a terminal event [1-5]. A challenge of MSMs is the combinatorial complexity in models as the number of states being modelled grows. This increases computation time, but also risks overfitting of some of the transition models within the MSM, especially those with fewer events/ observations.

Overfitting refers to when a model is too complex and captures noise/random fluctuations in the data rather than the true relationships between variables. In the case of MSMs, this usually manifests as too many parameters and overly complex model structures. One way overfitting can be diagnosed for MSMs is to assess the model's performance on data that was not used in the training process. Techniques such as cross-validation and comparison of prediction accuracy metrics such as the C-index between the training and test sets can aid in diagnosing overfitting in an MSM [6]. Overfitting in this context has been recognised as an issue in multi-state models [7][8].

One potential solution to this is to share parameters between transitions within the model [5]. For example, one could assume that certain groups of transitions share a common baseline and/or predictor effects. The properties of doing this in a prediction context are unknown. Likewise, in a prediction context, penalisation methods are often advocated to help minimise the risk of overfitting. Therefore, it is of interest to examine the impact of penalisation methods on this situation with or without sharing of parameter terms across the transitions. While the use of shared parameters across transitions to reduce the parameter space is well-documented, the concept of penalisation in this context is less well-studied.

Although sample size requirements for MSMs have not yet been developed, it is known that sample size is an important factor in determining a well-functioning model [9][10][11]. While penalisation may prove to be a useful tool to avoid overfitting, many papers have noted that penalisation is not a substitute for an adequate sample size [12][13][14]. Therefore, we will not go below the minimum sample size for a Cox model developed for each transition (in the absence of the MSM sample size criteria). We will explore the use of penalisation and parameter sharing as further protection against overfitting.

Definitions

We broadly define the terms relevant to this review:

Term	Definition
MLE	Maximum likelihood estimate
LASSO	Least absolute shrinkage and selection operator
Horseshoe prior	A type of Bayesian prior distribution, designed to encourage shrinkage of parameter estimates toward zero while allowing for the possibility of some parameters to remain large
Metropolis-Hasting method	A Markov Chain Monte Carlo method (MCMC) used for generating a sequence of random samples from a probability distribution
Burn-in period	The initial phase of the simulation where the Markov chain is allowed to converge to the target distribution
Bootstrap sampling	A sample made by randomly drawing observations with replacement from the original dataset. This creates new datasets of the same size as the original but with some observations omitted and others repeated

Table 1: A list of definitions used in the protocol.

Methodology

All analysis for this study will be conducted using R (version 4.3.1) in the RStudio environment under the MacOS Ventura operating system on a MacBook Pro (with Apple M1 Pro processor). Code for simulations and models will be provided on GitHub to ensure reproducibility.

Aims

This study aims to compare the predictive performance of current penalisation methodology for multi-state models. We will be considering the following methods of implementing the penalisation within prediction models:

- MLE no shrinkage
- MLE with uniform shrinkage factor applied to each transition
- LASSO penalised likelihood
- Bayesian approach with penalising prior to pull coefficients down towards an overall model-level value [15]

These methods will be applied to simulated data.

Data Simulation

We will consider two data generating models: first, a simple illness-death model, and second, a more complex 5-state model. This is done so that we can investigate the impact of the number of states (and transitions) on the predictive performance results. Each simulation will have n=200,000 individuals.

Simple Illness-Death Model

We will start by simulating data for a simple illness-death model (Figure 1). We will use the following Q-matrix:

$$Q = \begin{bmatrix} -0.02 & 0.015 & 0.005 \\ 0 & -0.01 & 0.01 \\ 0 & 0 & 0 \end{bmatrix},$$

using the *simmulti.msm* function from the *MSM* package to get a list of states and their transition times (starting in the healthy state at time 0) for each subject. The numbers in this Q-matrix were chosen in order to get sufficient progress through the states in the model while compromising on computation time.

The survival distributions for each transition were drawn from a hazard function of:

$$\exp[0.5 * age_i - 0.5 * gender_i + 0.5 * BMI_i] * \lambda_{j,k}$$

The baseline hazard for each transition ($\lambda_{j,k}$) was assumed to follow a Weibull distribution, where parameters were chosen based on the mean event time:

Table 1: Model parameters for each of the transitions in the three-state model

Transition	Shape, mean	Hazard function
1	2, 50	$h(t\|x) = 2\,exp(-18.02 + 0.34x_1 + 0.03\beta_2 + 0.41x_3)\,t$
2	25, 1000	$h(t\|x) = 25\,exp(-17.92 + 0.18x_1 + 0.02x_2 + 0.41x_3)\,t^{24}$
3	3.5, 50	$h(t\|x) = 3.5\,exp(-17.96 + 0.26x_1 + 0.02x_2 + 0.41x_3)\,t^{2.5}$

We will base the structure of the data on the structure of the multi-state oncology data dataset in R (onc3)[16].

5-State Model

The next model we will simulate is a 5-state model (Figure 2). We will use the following Q-matrix:

$$Q = \begin{bmatrix} -0.025 & 0.02 & 0 & 0 & 0.005 \\ 0 & -0.035 & 0.025 & 0 & 0.01 \\ 0 & 0 & -0.035 & 0.025 & 0.01 \\ 0 & 0 & 0 & -0.035 & 0.035 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

using the *simmulti.msm* function from the *MSM* package to get a list of states and their transition times (starting in the healthy state at time 0) for each subject. The numbers in this Q-matrix were

chosen in order to get sufficient progress through the states in the model while compromising on computation time.
The survival distributions for each transition were drawn from a hazard function of:

$$\exp[0.5 * age_i - 0.5 * gender_i + 0.5 * BMI_i] * \lambda_{j,k}$$

The baseline hazard for each transition ($\lambda_{j,k}$) was assumed to follow a Weibull distribution, where parameters were chosen based on the mean event time:

Table 2: Model parameters for each of the transitions in the five-state model

Transition	Shape, mean	Hazard function
1	2, 50	$h(t\|x) = 2\,exp(-18.02 + 0.34x_1 + 0.03\beta_2 + 0.41x_3)\,t$
2	25, 1000	$h(t\|x) = 25\,exp(-17.92 + 0.18x_1 + 0.02x_2 + 0.41x_3)\,t^{24}$
3	3.5, 50	$h(t\|x) = 3.5\,exp(-17.96 + 0.26x_1 + 0.02x_2 + 0.41x_3)\,t^{2.5}$
4	NTF	NTF
5	NTF	NTF
6	NTF	NTF
7	NTF	NTF

We will base the structure of the data on the structure of the multi-state oncology data dataset in R (onc3)[16] but extend it to include five states.

Penalisation Methods

Since we are using simulated survival data, a Cox proportional hazards model will be implemented using each of the penalisation methods in turn.

MLE No Shrinkage

To apply the MLE without shrinkage, a Weibull model will be fit to each transition in the model.

Uniform Shrinkage

To apply a uniform shrinkage factor to each transition, we will estimate the shrinkage factor post model fit using the heuristic shrinkage factor:

$$S_{VH} = 1 - \frac{p}{LR},$$

where p is the total number of predictor parameters for the full set of candidate predictors and LR is the likelihood ratio (chi-squared) statistic for the fitted model defined as:

$$LR = -2(lnL_{null} - lnL_{model}),$$

where lnL_{null} is the log-likelihood of a model with no predictors (e.g. intercept-only Weibull model), and lnL_{model} is the log-likelihood of the final model [17]. In the model, we will multiply each covariate by the shrinkage factor, allowing the shrinkage factor to be applied uniformly to both covariates across each transition in the model. This will be completed by multiplying each coefficient in each model by the shrinkage factor.

LASSO Penalised Likelihood

The LASSO penalised likelihood gives identical results to a Laplace prior with a Bayesian approach. We will be using the *brm* function from the *brms* package to implement this type of penalisation. We will specify priors for the intercept, standard deviation, and each covariate in the model. The models using the simulated data will not be complex, so we will only use 2 chains. We will start with 2000 (with a burn-in period of 1000 iterations) iterations, as this provides a good balance between computational efficiency. Using trace-plots and the Gelman-Rubin statistic, we will assess convergence; if these

diagnostic tools show that insufficient convergence has occurred, we will increase the number of iterations.

Bayesian Approach with Penalising Prior (Shared Parameters)

We have decided to use a horseshoe prior as described in Beesley (2020). A horseshoe prior is a continuous shrinkage prior, that strongly shrinks moderate and weak signals to zero, while still allowing very large signals. The prior takes the form:

$$f(\theta_k|\lambda_k,) = N(\alpha, \lambda_k^2\tau^2)$$
$$f(\lambda_k) = Cauchy^+(0,1)$$
$$f(\tau) = Cauchy^+(0,1),$$

where smaller τ implements a greater global shrinkage, θ_k are the log-hazard ratios for each of the k transition models, α is a non-zero constant (assuming that the 'shared' coefficient is non-zero across all transitions), and $Cauchy^+$ indicated the half-Cauchy distribution. τ, λ_k, and θ_k can be sampled in turn using the Metropolis-Hasting method. As we are applying the horseshoe prior to an MSM, we will define τ separately for each transition.

We will be using the *brm* function from the *brms* package to implement this type of penalisation. We will specify priors for the intercept, standard deviation, and each covariate in the model. The models using the simulated data will not be complex, so we will only use 2 chains. We will start with 2000 (with a burn-in period of 1000 iterations) iterations, as this provides a good balance between computational efficiency. Using trace-plots and the Gelman-Rubin statistic, we will assess convergence; if these diagnostic tools show that insufficient convergence has occurred, we will increase the number of iterations.

Packages to Use

Package	Use
mstate	Contains functions needed for data preparations, description, and hazard estimation for multi-state models [18].
tidyverse	A collection of packages for preparing, wrangling, and visualising data [19].
brms	Used for fitting a Bayesian Cox model with a horseshoe prior and a Laplace prior [20].

Table 4: A list of packages and their uses.

Performance Measures

- Computation time – As combinatorial complexity increases, so does complexity.
- Time-dependent ROC curves – To quantify discrimination.
- C-Index – To evaluate predictive performance.
- Calibration plots – Show any potential mismatches between observed and predicted probabilities in the data.
- Transition-specific calibration plots – To evaluate overfitting.

Validation

Internal Validation

For each iteration of the simulation, we will be generating a large ($n \geq 500{,}000$) sample size and will split this into two parts: a development subset and a validation subset. We will ensure that the validation subset is large ($n \geq 200{,}000$). For each penalisation method, we will fit a model using the validation subset using the same model specification as we used in the original analysis.

For each model fitted using the validation subset, we will calculate the following performance measures:

- Time-dependent ROC curves [21]
- C-Index [22]
- Calibration plots [23]
- Transition-specific calibration plots

These will be compared with the performance measures of the original model to check that the results are consistent between the development and validation subsets.

Author Contributions

CC and GM contributed to the conception and design of the study. All authors contributed to discussions around updates to the protocol.

CC wrote the protocol.

CC performed the data simulation and method implementation, iteratively updating the protocol where necessary.

Bibliography

[1] Heinz Freisling et al, 2020. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *National Library of Medicine,* 10 Jan. Harrison Reeder et al, 2022. Penalized Estimation of Frailty-Based Illness-Death Models for Semi-Competing Risks. *Biometrics.*

[2] Hein Putter et al, 2006. Estimation and Prediction in a Multi-State Model for Breast Cancer. *Biometrical Journal,* 16 June, 48(3), pp. 366-380.

[3] Niels Keiding et al, 2001. Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine,* 30 May, 20(12), pp. 1871-1885.

[4] Jeffrey Cannon et al, 2017. Rheumatic Heart Disease Severity, Progression and Outcomes: A Multi-State Model. *Journal of the American Heart Association,* 2 March.

[5] Holger Sennhenn-Reulen, 2016. Structured fusion lasso penalized multi-state models. *Statistics in Medicine,* 23 June, 35(25), p. 4637-4659.

[6] Jake Lever et al, 2016. Model selection and overfitting. *Nature Methods*, 30 August, 13, p703-704.

[8] Michael Crowther et al, 2017. Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. *Statistics in Medicine*, 05 September, 36(29), p4719-4742.

[9] Richard Riley et al, 2018. Minimum sample size for developing a multivariable prediction model: Part I - Continuous Outcomes. *Statistics in Medicine*, 22 October, 38(7), p1262-1275.

[10] Richard Riley et al, 2020. Calculating the sample size required for developing a clinical prediction model. *BMJ,* 18 March, 8238.

[11] Alexander Pate at al, 2023. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research*, 19 January, 32(3), p555-571.

[12] Ben Van Calster et al, 2020. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Statistical Methods in Medical Research,* 13 May.

[13] Richard Riley et al, 2021. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *Journal of Clinical Epidemiology,* April, 132, p88-96.

[14] Glen Martin et al, 2021. Developing clinical prediction models when adhering to minimum sample size recommendations: The importance of quantifying bootstrap variability in tuning parameters and predictive performance. *Statistical Methods in Medical Research,* December, 30(12), p2545-2561.

[15] Lauren Beesley et al, 2020. Bayesian variable selection and shrinkage strategies in a complicated modelling setting with missing data: A case study using multistate models. *Statistical Modelling*, 21(1-2), p11-29.

[16] Devin Incerti, 2022. onc3: Multi-state oncology data for 3-state model. Available at: https://rdrr.io/cran/hesim/man/onc3.html
[Accessed 19 11 2023]

[17] Richard Riley et al, 2018. Minimum sample size for developing a multivariable prediction model: PART II - binary. *Statistics in Medicine,* 24 Oct, 38(7), p1276–1296.

[18] Hein Putter et al, 2021. *mstate: Data Preparation, Estimation and Prediction in Multi-State Models.* [Online]
Available at: https://cran.r-project.org/web/packages/mstate/index.html
[Accessed 9 11 2023].

[19] Hadley Wickham, 2023. *tidyverse: Easily Install and Load the 'Tidyverse'.* [Online]
Available at: https://cran.r-project.org/web/packages/tidyverse/index.html
[Accessed 9 11 2023].

[20] Paul Christian Bürkner, 2023. *Package 'brms'.* [Online]
Available at: https://cran.r-project.org/web/packages/brms/brms.pdf
[Accessed 9 11 2023].
[21] Frank Harrell et al, 1996. Multivariable Prognostic Models: Issue in Developing Models, Evaluating Assumptions and Adequacy, and Measure and Reducing Errors. *Statistics in Medicine* 15(4), p361-387.
[22] Patrick Heagerty et al, 2005. Survival Model Predictive Accuracy and ROC Curves. *Biometrics* 61(1), p92-105.
[23] Ben Van Calster et al, 2019. Calibration: the Achilles heel of predictive analytics. *BMC Medicine* 17(230).

Figures

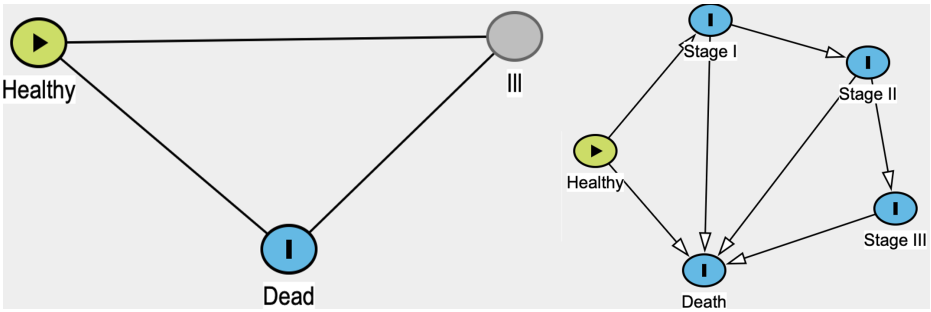

Figure 1: The simple illness-death model (left), and the five-state model (right).