

Multi-State Clinical Prediction Models in Renal Replacement Therapy

A Thesis

Presented to

The Division of Division of Informatics, Imaging and Data Science

University of Manchester

In Partial Fulfillment

of the Requirements for the Degree

PhD Medicine

Michael Andrew Barrowman

March 2020

Approved for the Division
(School of Health Sciences)

Dr. Matthew Sperrin, Prof. Niels Peek, Dr. Glen Martin, Dr. Mark Lambie

Acknowledgements

I would like to thank my...

Table of Contents

Introduction	1
0.1 Subsectioning	1
Chapter 1: Literature Report	3
1.1 Introduction	3
1.2 Clinical Prediction Models	3
1.2.1 Fundamental Prognosis Research	3
1.2.2 Prognostic Factor Research	4
1.2.3 Prognostic Model Research	4
1.2.4 Stratified Medicine	4
1.2.5 Examples	4
1.3 Competing Risks & Multi-State Models	4
Chapter 2: The Application of Multi-State Methods to Develop Clinical Prediction Models Designed for Clinical Use - A Scoping Review	5
2.1 Introduction	5
2.2 Methods	6
2.2.1 Scope of Review	6
2.2.2 Initial Search Strategy	7
2.2.3 Filtering	8
2.2.4 Data Extraction	9
2.2.5 Reporting	9
Chapter 3: How unmeasured confounding in a competing risks setting can affect treatment effect estimates in observational studies . . .	11
3.1 Background	11
3.2 Methods	12
References	15

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Introduction

Overall introduction to the Thesis. This doesn't seem to want to show up.

Why is this?

0.1 Subsectioing

Testing

Chapter 1

Literature Report

This is the first chapter of my thesis and will include a brief summary of what the current literature looks like. It will be split into sections and subsections as specified in my ToDo List

1.1 Introduction

lorem ipsum blah blah blah

1.2 Clinical Prediction Models

The idea of prognosis dates back to ancient Greece with the work of Hippocrates [**Cite: Hippocrates**] and is derived from the Greek for “know before” meaning to forecast the future. Within the sphere of healthcare...

Prognosis research can be broken down into four main categories (with three subcategories [1]):

- Type I: Fundamental prognosis research [**Cite: LR:1**]
- Type II: Prognostic factor research[**Cite: LR:2**]
- Type III: Prognostic model research[**Cite: LR:3**]
 - Model development[**Cite: LR:34**]
 - Model validation[**Cite: LR:35**]
 - Model impact evaluation[**Cite: LR:36**]
- Type IV: Stratified Medicine [**Cite: LR:4**]

1.2.1 Fundamental Prognosis Research

What is it?

1.2.2 Prognostic Factor Research

The aim of prognostic factor research (Type II) is to discover which factors are associated with disease progression. This allows for the general attribution of relationships between predictors and clinical outcomes.

Predictive factor research can give researchers and clinicians an idea of which patient factors are important when assessing a disease. It is vital to the development of clinical predictive models as without an idea of what covariates *can* affect an outcome, we cannot figure out which variables *will* affect the outcome. For example, [xxxx] demonstrated that [xxxx] is correlated with [xxxx], which subsequently used as a covariate in the development of the [xxxx] model.

1.2.3 Prognostic Model Research

Model Development

Model Validation

Impact Evaluation

1.2.4 Stratified Medicine

1.2.5 Examples

1.3 Competing Risks & Multi-State Models

Chapter 2

The Application of Multi-State Methods to Develop Clinical Prediction Models Designed for Clinical Use - A Scoping Review

2.1 Introduction

eHealthcare is moving towards a more data-driven approach to decision making, exploiting the variety of data sources collected as part of routine care [1]. This increases efficiency, which is becoming increasingly vital as patients are living longer and requiring more care, while budgets are being reduced [2], [3]. Correspondingly, there has been a shift towards primary prevention, rather than purely treating disease as it arises [4] therefore clinical prediction models (CPMs) are more relevant than ever before [5].

Prognostic CPMs (those that predict the future) allow end-users to estimate an individual's probability/risk of experiencing an outcome of interest within a certain timeframe. CPMs are algorithms that relate a set of prognostic factors to the risk of a chosen outcome [6], often using multivariable regression. They can provide predictions of the future course of an illness and provide evidence for the commencement of medical interventions [7].

Along with this overall increase in importance, different methods of producing CPMs are also being used, and each makes different assumptions, and models at different levels of granularity. One of these methods is the Multi-State Model (MSM), an extension to traditional survival analysis wherein patients exist in one of many distinct states at any given time and can transition between them (these individual transitions are akin to that of traditional survival analysis) [8]. A subset of MSMs is that of a Competing Risks model, where patients can only move from a single initial state to many absorbing states without any intermediate or transient states. A huge advantage of Multi-State CPMs, and indeed, Competing Risks CPMs, is that they can provide predictions for multiple outcomes with MSMs going further by allowing

the prediction of multiple pathways to that outcome, whereas traditionally developed models only provide predictions for a single end-point.

However, little is known about how widely these types of models are implemented in clinically relevant prognostic research. Therefore, we here aim to document a scoping review protocol that will intend to uncover any prediction models using MSMs that have been developed for clinical use. As part of the process of this investigation, we will also document how many CPMs account for Competing Risks alone. We define a scoping review as described by Arksey and O'Malley [9], which is similar to a systematic review, but with less formal outline for the analysis and synthesis of literature [10]. By assessing how MSMs have currently been applied in this field, we aim to describe the landscape of their current use, the context in which they are being used and discuss ways in which their use, application and uptake can be improved. To the best of our knowledge, a review such as this for Multi-State Models has never been performed.

2.2 Methods

$$A = \pi r^2$$

2.2.1 Scope of Review

This review will cover articles related to the development of Multi-State Clinical Prediction models designed for clinical use. It will not include models that were developed solely for demonstrations of novel methodological improvements in the field of clinical prediction modelling and/or multi-state modelling. Article inclusion will be based on the screening of the article text and interpretation of its aims, primary distinction will be made on whether an existing dataset is used as a core part of the article or as a subsidiary example. It will include articles that validate previously developed models and those that review existing models, only so far as to use them to find the original development article (a method known as Snowballing).

As this analysis will follow the style of a scoping review; the final paper will adhere to the PRISMA-ScR guidelines [11], which were set out to extend the traditional PRISMA guidelines to a Scoping Review setting.

Models which focus only on a competing risks scenario (whether directly or simply adjusting for competing risks) will not be analysed in detail, however to avoid missing possible Multi-State Models, we will only omit these at the final stage of screening (See below). This will also allow for a brief description of how many CR models exist compared to the MSM models to be analysed in detail in this review.

As per the definitions set out by the PROGRESS research group, prognostic research is split into four overarching themes/types:

- Type I - Fundamental Prognosis Research [12]
- Type II - Prognostic Factor Research [13]
- Type III - Prognostic Model Research [14]
- Type IV - Stratified Medicine Research [15]

As such, we will be focusing on papers of Type III [14]. Articles related to the other types of prognostic research often develop a model within their work, but since the intent of these papers is to investigate overall outcomes, effects of an individual factor or interactive effects of treatments in individuals, they are considered disjoint from CPM development and so they will not be included in our analysis.

2.2.2 Initial Search Strategy

Search Terms

To ensure we cover as much of the medical literature as possible, we will use the Ovid search engine to search two databases:

- EMBASE (1974 to 2018 December 31)
- Ovid MEDLINE and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions 1946 to December 31, 2018

We will use a standard set of terms designed by Ingui & Rogers [16] and added to by Geersing et al [17] used for searching for clinical prediction related literature. We will also extend this by including search terms relating to time-to-event outcomes and/or survival analysis that were defined by the authors, and which aim to broaden our search (see table 1). This will be combined by a set of search terms designed to filter for MSMs and/or CRs.

These novel MSM/CR terms include “fine adj2 gray” to include papers which use the Fine & Gray subdistribution proportional hazard method [18]. It will also include “semimarkov or semi markov” to include articles which specify that the model adopts a semi-Markov perspective, which is common amongst MSMs [8]. However, we chose not to include the term “markov” alone as it is considered to be too unspecific to be of use (a la search for “model” alone when finding clinical prediction models). The full search details can be found in table 2.

We believe that the broadness of our search terms allows for high sensitivity in our results and will therefore provide a larger and more comprehensive pool of papers than using a more specific set of search terms.

[Insert Table from paper]

Validation set of articles

To ensure that our search strategy is satisfactory, we will compare our results to a set of Validation papers. These are papers that we are already aware of that satisfy our inclusion/exclusion criteria and which therefore should be included in our analysis. We will compare the results of our initial search with this set of papers to ensure that all of the Validation set appear in our results. If they do not, then we will adjust our search strategy iteratively increasing sensitivity and improving the reach of our search until all Validation papers are included. The set of Validation papers is as follows:

- *Estimation and Prediction in a Multi-State Model for Breast Cancer*, Putter et al, 2006 [20]

- *A Multi-State Model to Predict Heart Failure Hospitalizations and All-Cause Mortality in Outpatients With Heart Failure With Reduced Ejection Fraction: Model Derivation and External Validation*, Upshaw et al, 2016 [21]
- *Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate*, Grams et al, 2018 [22]
- *Estimating transition probability of different states of type 2 diabetes and its associated factors using Markov model*, Nazari et al, 2018 [23]
- *Advantages of a multi-state approach in surgical research: how intermediate events and risk factor profile affect the prognosis of a patient with locally advanced rectal cancer*, Manzini et al, 2018 [24]

2.2.3 Filtering

Once the initial set of articles has been found, these will be filtered at various degrees of granularity to focus on papers which are included in the scope of our review as per our inclusion/exclusion criteria. We will also define which papers will be used only for the snowballing process, but will not be used as part of our analysis.

Inclusion/Exclusion Criteria

Inclusion

- Type III Prognostic Study Papers (i.e. those developing a clinical prediction model) [14]
- Papers which use a Multi-State Model framework to provide individual level patient predictions

Exclusion

- Papers that develop overall population level predictions (Type I)
- Papers focused on identification of prognostic factors (Type II)
- Papers that investigate stratified medicine (Type IV)
- Papers that only develop Competing Risks models
- Papers designed to describe methodological models with or without clinical application used only for an example

Stages

The filtering of the results will be performed in three stages: 1. Title (MB) 2. Abstract (MB with 20% replication by DJ) 3. Full Paper (MB with 20% replication by DJ)

Filtering will begin with an initial check through all titles to assess whether it is believed that the paper may be relevant to the review. This will help to omit a large amount of papers that were incorrectly returned by the broad search strategy. To ensure the review remains as sensitive as possible, only papers where it is abundantly clear that they violate an inclusion/exclusion criteria will be removed at this stage.

A second filter will be performed on the abstracts of the remaining articles and removed papers will be classified by the reason for their omission. To allow for

faster data extraction, a final glancing filter will also be performed over the full papers to again reduce the numbers of collated papers in the final review and reduce the likelihood of removing papers at the analysis stage. To ensure robustness of this filtering, both of these stages will be replicated by a second reviewer (DJ) in a randomly selected 20% of the abstracts and papers and differences will be discussed internally. At this point, models focusing solely on competing risks (i.e. those without a transient state) will be filtered out.

2.2.4 Data Extraction

To study the use of Multi-State Clinical Prediction Models from a quantitative perspective, certain vital data points will be extracted from the extant models. These measurements can be grouped as to what element of the prediction model they are evaluating:

- * Clinically Relevant points
- * Number of patients
- * Clinical setting (i.e. primary vs secondary care, geographic setting)
- * Field of study (e.g. cardiovascular, renal, etc.)
- * Summary of patient demographics (i.e. inclusion/exclusion criteria)
- * Outcomes being predicted
- * Multi-State Model details
- * Number of States and what they are
- * Shape/Structure of the model (i.e. how patients can transition between states)
- * How were relevant variables chosen?
- * Transition assumptions (e.g. parametric vs non-parametric, PH assumption, etc. . .)
- * Stated justification for, and reported benefits of an MSM versus traditional methods.
- * Predictive Ability
- * Timeframe (e.g. single time point(s), continuous time prediction, dynamic prediction, etc. . .)
- * What validation was performed (None vs. Internal (bootstrap, CV, etc.) vs. External)
- * Comparisons to current guidelines
- * Assessment of Bias of their model (using PROBAST)
- * Utilisation of the TRIPOD Guidelines (e.g. Was it referenced? Was it adhered to?)
- * Prominence information
- * Number of citations (although not clinically relevant, it is relevant to understanding the model's utilisation)
- * Year of publication (again, not clinically relevant, but useful to spot any time trends in prominence and/or quality)

The data extracted at this stage will be checked by DJ in 20% of the papers to confirm results for the analysis

2.2.5 Reporting

The search and filtering strategy will be depicted with a modified PRISMA flow diagram [30], which includes papers found by Snowballing and how they are included in the filtration process, see figure 1.

[Add in PRISMA]

A table of the extracted information will be included with the paper, depending on the number of results, this may be supplementary material. This information will also be summarised and analysed both quantitatively and qualitatively. For example, as the Illness-Death model [8] is simple and common amongst multi-state models, we will count how many of the MSCPMs use this structure as well as the other most common structures used. Any direct comparisons that can be made between predictions of this type (i.e. from the same field with the same outcomes) will be described.

Chapter 3

How unmeasured confounding in a competing risks setting can affect treatment effect estimates in observational studies

3.1 Background

Well-designed observation studies permit researchers to assess treatment effects when randomisation is not feasible. This may be due to cost, suspected non-equipoise treatments or any number of other reasons [1]. While observational studies minimise these issues by being cheaper to run and avoiding randomisation (which, although unknown at the time, may prescribe patients to worse treatments), they are potentially subject to issues such as unmeasured confounding and increased possibility of competing risks (where multiple clinically relevant events occur). Although these issues can arise in any study, Randomised Controlled Trials (RCTs) attempt to mitigate these effects by using randomisation of treatment and strict inclusion/exclusion criteria. However, the estimated treatment effects from RCTs are of potentially limited generalisability, accessibility and implementability [2].

A confounder is a variable that is a common cause of both treatment and outcome. For example, a patient with a high Body Mass Index (BMI) is more likely to be prescribed statins [3], but are also more likely to suffer a cardiovascular event. These treatment decisions can be affected by variables that are not routinely collected (such as childhood socio-economic status or the severity of a comorbidity [4]). Therefore, if these variables are omitted from (or unavailable for) the analysis of treatment effects in observational studies, then they can bias inferences [5]. As well as having a direct effect on the event-of-interest, confounders (along with other covariates) can also have further reaching effects on a patient's health by changing the chances of having a competing event. Patients who are more likely to have a competing event are less likely to have an event-of-interest, which can affect inferences from studies ignoring the competing event. In the above BMI example, a high BMI can also increase a

patient’s likelihood of developing (and thus dying from) cancer [6].

The issue of confounding in observational studies has been researched previously [7,8,9], where it has been consistently shown that unmeasured confounding is likely to occur within these natural datasets and that there is poor reporting of this, even after the introduction of the The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Guidelines [10, 11]. Hence, it is widely recognised that sensitivity analyses are vital within the observational setting [12]. However these previous studies do not extend this work into a competing risk setting, meaning research in this space is lacking [13], particularly where the presence of a competing event can affect the rate of occurrence of the event-of-interest. These issues will commonly occur in elderly and comorbid patients where treatment decisions are more complex. As the elderly population grows, the clinical community needs to understand the optimal way to treat patients with complex conditions; here, causal relationships between treatment and outcome need to account for competing events appropriately.

The most common way of analysing data that contains competing events is using a cause specific perspective, as in the Cox methodology [14], where competing events are considered as censoring events and analysis focuses solely on the event-of-interest. The alternative is to assume a subdistributional perspective, as in the Fine & Gray methodology [15], where patients who have competing events remain in the risk set forever.

The aim of this paper is to study the bias induced by the presence of unmeasured confounding on treatment effect estimates in the competing risks framework. We investigated how unmeasured confounding affects the apparent effect of treatment under the Fine & Gray and the Cox methodologies and how these estimates differ from their true value. To accomplish this, we used simulations to generate synthetic time-to-event-data and then model under both perspectives. Both the Cox and Fine & Gray models provide hazard ratios to describe the effects of a covariate. A binary covariate will represent a treatment and the coefficients found by the model will be the estimate of interest.

3.2 Methods

We considered a simulation scenario in which our population can experience two events; one of which is the event-of-interest (Event 1), the other is a competing event (Event 2). We model a single unmeasured confounding covariate, $U \sim N(0, 1)$ and a binary treatment indicator, Z . We varied how much U and Z affect the probability distribution of the two events as well as how they are correlated. For example, Z could represent whether a patient is prescribed statins, U could be their BMI, the event-of-interest could be cardiovascular disease related mortality and a competing event could be cancer-related mortality. We followed best practice for conducting and reporting simulations studies [16].

The data-generating mechanism defined two cause-specific hazard functions (one for each event), where the baseline hazard for event 1 was k times that of event 2, see Fig. 1. We assumed a baseline hazard that was either constant (exponential

distributed failure times), linearly increasing (Weibull distributed failure times) or biologically plausible [17]. The hazards used were thus:

$$\lambda_1(t|U, Z) = ke^{\beta_1 U + \gamma_1 Z} \lambda_0(t) \quad (3.1)$$

$$\lambda_2(t|U, Z) = ke^{\beta_2 U + \gamma_2 Z} \lambda_0(t) \quad (3.2)$$

$$\lambda_0(t) \begin{cases} 1 & \text{Exponential} \\ 2t & \text{Weibull} \\ \exp -18 + 7.3t - 11.5t^{0.5} \log(t) + 9.5t^{0.5} & \text{Plausible} \end{cases} \quad (3.3)$$

In the above equations, β and γ are the effects of the confounding covariate and the treatment effect respectively with the subscripts representing which event they are affecting. These two hazard functions entirely describe how a population will behave [18].

[Insert Figure 1]

We simulated populations of 10,000 patients to ensure small confidence intervals around our treatment effect estimates in each simulation. Each simulated population had a distinct value for β and γ . In order to simulate the confounding of U and Z , we generated these values such that $\text{Corr}(U, Z) = \rho$ and $\text{Pr}(Z = 1) = \pi$ [19]. Population end times and type of event were generated using the relevant hazard functions. The full process for the simulations can be found in Additional file 1. Due to the methods used to generate the populations, the possible values for ρ are bounded by the choice of π such that when $\pi = 0.5$, $|\rho| \leq 0.797$ and when $\pi = 0.1$ (or $\pi = 0.9$), $|\rho| \leq 0.57$. The relationship between the parameters can be seen in the Directed Acyclic Graph (DAG) shown in Fig. 2, where T is the event time and δ is the event type indicator (1 for event-of-interest and 2 for competing event).

References

- [1] R. D. Riley, D. van der Windt, P. Croft, and K. G. M. Moons, *Prognosis Research in Healthcare: Concepts, Methods, and Impact*, First. Oxford University Press, 2019.