Markovian multi-state models and an application to Alzheimer's Disease

Broer, Lana

S.D.L. Broer

Markovian multi-state models and an application to Alzheimer's Disease

Bachelor thesis

July 21st, 2023

Thesis supervisor: Prof.dr. M. Fiocco

Leiden University
Mathematical Institute

Contents

1 Introduction

Multi-state models (MSMs) are a popular approach in the handling of longitudinal data. With this type of model a process with several outcomes, or multiple intermediate stages, can be represented. Thereby, the multi-state model is popular in biomedical research, where it can be used to model and predict how patients move through stages of a disease or treatment. For example, MSMs may be used to model how patients progress after removal of a cancerous tumour, from post-operative remission to local recurrence and/or metastasis and eventually death [35]. Multi-state models can also be used to model the progression of dementia, as has previously been done by [6, 20, 25, 39, 47, 48, 49]. This latter example will be the main application of this thesis.

With multi-state models, it is possible to estimate the probability to move from one stage to another. Many approaches have been described with which these probabilities, the so-called transition probabilities, can be estimated. These approaches make a distinction between the (semi-)Markovian and non-Markovian models, and can use non-parametric, semi-parametric and parametric estimation methods. The non-parametric estimators, such as the Aalen-Johansen estimator, do not allow for incorporation of covariate effects. Therefore, other approaches are needed if there is wish to consider such effects. A possibility to do so, is to firstly estimate the covariate effects using the popular Cox regression model, as described by Putter, Fiocco and Geskus (2006) [35], after which the covariate effects can be used in the estimation of the transition probabilities.

In the studies performed on the progression of dementia using multi-state models, several approaches have been taken to estimate covariate effects on the transition probabilities. Salazar et al. (2007) and Yu et al. (2009) modeled the progression to dementia using two intermediate stages between the starting and demented state [39, 49]. In their models, they used two absorbing states - dementia and death -, and used a polytomous logistic regression model with shared random effects to estimate the effect of variables on the transitions. Kryscio et al. (2006) similarly used polytomous logistic regression models to model the covariate effects [25]. Wang et al. (2023) proposed a three-step estimation method to use fixed and random effects to estimate the covariate effects on the transition probabilities, effectively using a generalized linear mixed model approach [47]. Commenges and Joly (2004) created a model to evaluate the interaction between dementia and institutionalization using a penalized likelihood approach [6]. This approach is elaborated further in the study described by Joly et al. (2009) [20].

In these previously performed studies, Cox regression has not yet been used to estimate the covariate effects for the transition probabilities. In the study by Wei, Xu and Kryscio (2014), it was shown that the maximum likelihood estimates of the Weibull and Cox models were stable under violations of their model assumptions in the estimation of covariate effects on each transition [48]. In our research, we aim to use the Cox proportional hazards regression model to estimate the progression to dementia, including death as an absorbing but not competing state, using the Cox regression model to estimate covariate effects, which will be used in the estimation of the transition probabilities.

The aims of this thesis are to (i) give an overview of the theory behind multi-state models and present a way to estimate them, and (ii) show an application of multi-state models using real-life data from the Alzheimer's Disease Neuroimaging Initiative*, focusing on modelling the progression to dementia using a Cox proportional hazards regression model and comparing model outcomes on two separate time scales. To do so, first a general introduction to the most important concepts in survival analysis will be given, necessary for understanding multi-state models, followed by the theoretical framework of multi-state models, and ending with the application using data on Alzheimer's Disease.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf [19].

2 Basic concepts

This chapter deals with the basic concepts that underlie survival analysis.

Survival analysis is the branch of statistics that deals with time-to-event data. Time-to-event data consists of longitudinal data in which patients are followed over time from a set starting point until development of the event of interest, or until the end of follow-up, whichever comes first [27]. A multitude of options exist for events of interest, with death being the most classical example. Other examples are relapse of breast cancer after chemotherapy, or being symptom-free after an infection with COVID-19. As such, events of interest may describe both negative (death, relapse) or positive (being symptom-free) events.

In survival analysis, the time at which the event of interest occurs is of specific interest. Thereby, analyses within this field focus on making statements about the time it takes for an event to occur. To do so, it is necessary to record the status of the participant with respect to the event of interest at each time point. As such, information about the time and status are oftentimes recorded together and combined in the pair of random variables (X, δ). Here, the time at which the event of interest takes place is captured by the random variable X and δ indicates the event having occurred/not occurred (1/0) at time x [22, Ch. 3.2; 23, Ch. 1.1.1]. It should be noted that x_0, called the baseline, is participant specific and is defined as the moment from which the time scale for that individual started. Additionally, models may be fit on several time scales. On the so-called study scale, time counts from the moment of inclusion in the study, whereas on the age-scale, time is counted from birth. Using a different time scale, changes the interpretation of the model.

2.1 Censoring

Within survival analysis, censoring is a common occurrence. Censoring occurs when the event of interest happens outside of the follow-up period [22, Ch. 3.1]. Evidently, some type of censoring is nearly inevitable unless patients are measured continuously for the event of interest. Studies without censoring can be imagined, such as a study conducted on patients in an intensive care unit on continuous monitoring devices, however such studies are rare. As such, most, if not all, methods in survival analysis make assumptions about the censoring mechanisms or are capable of handling censored data. Violation of this assumption will lead to biased estimates [11, 35].

Several types of censoring are distinguished [22, Ch. 3; 23, Ch. 1.1]:

- **Right censoring**: when the event of interest happens *after* the follow-up period ends, the observation is said to be right censored, i.e. $x_{last} < x_{event}$. An example of right censored data is a study where patients are followed until death, but an individual drops out due to relocating to another country. Right censoring is the most common type of censoring.

- **Left censoring**: when the event of interest happens *before* the follow-up period starts, the observation is said to be left censored, i.e. $x_0 > x_{event}$. An example of left censoring could be a study that is conducted on the age at which young adults start smoking, but some individuals had already started smoking before being included in the study.

- **Interval censoring**: when the event of interest happens *between* two measurements, the observation is said to be interval censored, i.e. $x_i < x_{event} < x_{i+1}$. For example, a study is conducted with primary interest in relapse of breast cancer after chemotherapy, where yearly check-ups are performed. When an individual is diagnosed with relapse at the third check-up, the event of interest occurred between the second and third year, but the exact moment is unknown.

Since either the censoring time or the event time is observed, but never both, we introduce the random variable $T = \min(X, C_r)$, which is the minimum of the event time X and the right censored time C_r. In other words, T describes the observed time. Similar random variables could be introduced for the left and interval censored data.

Apart from the types of censoring, a distinction is made between two censoring mechanisms: informative and non-informative censoring. Both censoring mechanisms may occur in each censoring type. Informative censoring describes the situation in which the censoring mechanism is related to the event of interest. For example, in a study where HIV-patients are followed from the moment of diagnosis until development of AIDS, but patients who experience more symptoms are more likely to drop out, the censoring mechanisms is informative. Contrarily, non-informative censoring indicates a censoring mechanism in which censoring is unrelated to the event of interest. Generally, it is assumed that censoring is non-informative. As such, the presence of informative censoring in the data may lead to biased estimates [22, Ch. 4.2].

2.2 Truncation

Truncation is a concept similar to censoring and is similarly commonplace in survival analysis. However, unlike censoring, truncation is generally described as being a property of the data. Truncation happens when observations are included only when the event of interest is experienced in the observational window. Several types of truncation are distinguished, with the types of truncation being similar to the types of censoring. However, the crucial difference is that censored observations are included in the dataset, whereas truncated events are not a part of the data. In this way, truncation is more invisible than censored data and requires consideration in the study design. Moreover, since no information is available on individuals that do not experience the event in the observational window, only conditional probabilities can be estimated in the case of truncated data [22, Ch. 3.4].

Several types of truncation are recognized:

- **Right truncation**: when all individuals who experience the event of interest *after* the observational window are excluded, the data is said to be right truncated. An example of right truncation is a study on mortality based on death records.

- **Left truncation**: when all individuals who experience the event of interest *before* the observational window are excluded, the data is said to be left truncated. For example, a study that is performed on the time to death in a

nursing home population. Thereby individuals that die before being able to go into a nursing home cannot participate in the study.

Left truncation, also called late entry, is the most common type of truncation. Note that studies with left truncation may suffer from late entry bias if left truncation is not accounted for.

- **Double truncation**: when all individuals who experience the event of interest *before or after* the observational window are excluded, the data is said to be double truncated. For example, a study on mortality using death records from a nursing home population would contain double truncation.

2.3 Survival function

The survival, which is defined as the event-free period, is quantified through the survival function. The survival function describes the probability to survive past a specific time point $x \geq 0$ [22, Ch. 2.2 and Ch. 4.2]:

$$S(x) = \mathbb{P}(X > x).$$

Commonly, X is measured on a continuous scale, however, discrete realizations of X are possible. This happens, for example, when events are measured on intervals, rather than at specific time points.

The survival function can be estimated using the Kaplan-Meier estimator, also called the product-limit estimator [21]. It assumes the censoring and event time distributions to be independent from each other. The step-function that is estimated with the Kaplan-Meier evaluates the number of events (d_i) and the number of patients still in follow-up (Y_i) at each time point (x_i) preceding the time at which the estimator is evaluated (x) to calculate the survival:

$$\hat{S}_{KM}(x) = \prod_{x_i \leq x} (1 - \frac{d_i}{Y_i}).$$

The standard error (SE) of the Kaplan-Meier can be estimated using Greenwoods' formula [23, Ch. 1.1.2]:

$$\text{se}(\hat{S}_{KM}(x)) = \sqrt{\sum_{x_i \leq x} \frac{d_i}{Y_i(Y_i - d_i)}}.$$

2.4 Hazard function

Alternatively to the survival function, the rate at which the event of interest is experienced can be expressed through the hazard. The hazard, also called the failure rate, is the probability of experiencing the event of interest instantaneously. The hazard is given by ([35])

$$h(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x | X \geq x)}{\Delta x}.$$

Furthermore, the hazard at each time point can be aggregated into the cumulative hazard, which describes the total hazard at a given time x. The cumulative hazard is ([22, Ch. 2.3])

$$H(x) = \int_0^x h(u)du.$$

When X is measured continuously and $S(x)$ is differentiable, we can rewrite the hazard and cumulative hazard functions as follows ([22, Ch. 2.3; 23, Ch. 1]),

$$h(x) = \lim_{\Delta x \downarrow 0} \frac{S(x + \Delta x)}{S(x)} = -\frac{S'(x)}{S(x)} = -\frac{d \ln(S(x))}{dx};$$

$$H(x) = -\ln(S(x)).$$

Thereby, we can also rewrite the survival function in terms of the hazard,

$$S(x) = e^{-H(x)} = e^{-\int_0^x h(u)du}.$$

Clearly, the (cumulative) hazard and survival are inseparably related and can always be deduced from each other.

Similar to the survival function, the hazard can be estimated using the Kaplan-Meier estimator by substituting $\hat{S}_{KM}(x)$ in the hazard function. Alternatively, the Nelson-Aalen estimator can be used to estimate the hazard function [1]. It assumes the censoring and event time distributions to be independent. The Nelson-Aalen estimator uses the same concepts as the Kaplan-Meier estimator, namely the number of events (d_i) and the number of patients still in follow-up (Y_i) at each time point (x_i) preceding the time at which the estimator is evaluated (x):

$$\hat{H}_{NA}(x) = \sum_{x_i \leq x} \frac{d_i}{Y_i}. \tag{1}$$

It's standard error is given by ([23, Ch. 1.1.3])

$$\text{se}(\hat{H}_{NA}(x)) = \sqrt{\sum_{x_i \leq x} \frac{d_i}{Y_i^2}}.$$

In some situations, it can be useful to compare the hazards of two groups. For example, in a study on smoking, interest may be in if women are more likely to start smoking than men. This variation in hazard can be quantified through the hazard ratio (HR), which is calculated by dividing the hazards of the two groups. The HR expresses how much more likely the group in the numerator is to experience the outcome, compared to the group in the denominator.

2.5 Likelihood

The likelihood function, which outcome is used as a measure of plausibility, is a crucial concept in statistical inference. Since censoring and truncation complicate the derivation of the likelihood, an adaptation of (or addition to) the general likelihood

function exists in the field of survival analysis. All information discussed below can be found in [22, Ch. 3.5].

In case of censoring, we can construct the likelihood as follows. Let $f(x)$, $S(C_r)$, $1 - S(C_l)$ and $S(L) - S(R)$ represent the survival times of complete cases, right-censored observations, left-censored observations and interval-censored observations, respectively. Let D, R, L, I be the sets of complete, right-censored, left-censored and interval-censored observations, respectively. Then, the likelihood is as follows:

$$L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_r) \prod_{i \in L} (1 - S(C_l)) \prod_{i \in I} [S(L_i) - S(R_i)].$$

When only right censoring is present in the data, the likelihood can be simplified by using the pair of random variables (T, δ) as previously described, where instead of the event time X the observed time T is used:

$$L = \prod_{i=1}^{n} \mathbb{P}(t_i, \delta_i) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}.$$

The likelihood can also be rewritten in terms of the (cumulative) hazard by using that $f(t_i) = h(t_i)S(t_i)$:

$$L = \prod_{i=1}^{n} h(t_i)^{\delta_i} e^{-H(t_i)}.$$

2.6 Types of models

Survival estimations come in three flavours: parametric, semi-parametric and non-parametric. Parametric models use probability distributions in their estimation, imposing strict assumptions on the structure of the data. Contrarily, non-parametric models make no assumptions about a possible underlying distribution. Semi-parametric models, acting as a middle ground, introduce an assumption about the structure of some parts of the data, but not on the overall data.

Parametric models are often used theoretically, since the imposed structure on the data makes calculations efficient. The exponential, Weibull and lognormal distributions are commonly used distributions in paramateric survival models. In practice, semi-parametric and non-parametric models are widely used due to the flexibility they allow in the structure of the data. For example, the Kaplan-Meier and Nelson-Aalen are non-parametric estimators. However, a disadvantage of non-parametric estimators is that they do not allow for the use of covariates. As such, semi-parametric models are used to introduce covariates into non-parametric models. The most popular semi-parametric model in survival analysis is the Cox proportional hazards regression model, which imposes a parametric distribution on the covariates, but not on the hazard.

The Cox proportional hazards regression model defines a conditional hazard on the covariates of interest ([23, Ch. 1.2]):

$$h(x|Z) = h_0(x)e^{Z^T\beta},$$

where h_0 is the baseline hazard at time x, Z is a column vector containing the covariate information of an individual and β is the vector of regression coefficients

estimating the effect of each covariate on the hazard. Here, the baseline hazard $h_0(x)$ is the hazard given an individual with all covariates in the reference state, i.e. all covariates are equal to zero. Note that instead of a column vector Z, a design matrix may be used, which results in a vector containing the hazard for each individual. A useful feature of Cox regression is that the exponent of the estimated regression coefficients are equal to the HR's. This makes the model outcomes easily interpretable. The Cox proportional hazards regression model assumes proportional hazards, i.e. the HR of any two individuals is constant over time.

2.7 Competing risks

In survival analysis, it may happen that several events compete as an outcome. For example, when interest is in occurrence of relapse after chemotherapy, patients may die before relapsing. In this case, we call death a competing event for experiencing a relapse. In general, when we have events that prevent our event of interest from happening, we call these competing events (see Figure 1).

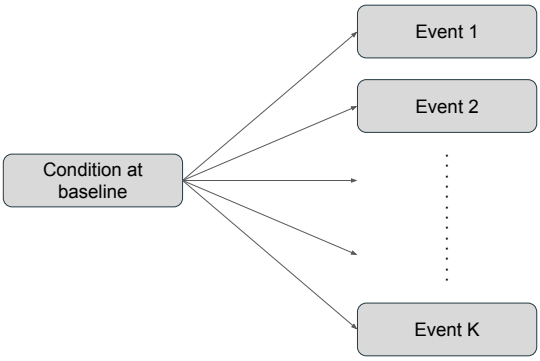


Figure 1: A general model with competing events.

Cause-specific hazard functions may be defined for such circumstances, which are analogous to the marginal hazard function discussed in Section 2.4. The cause-specific hazard function is defined specifically for cause $k \in K$, with K being the set of all possible events and D the random variable indicating the event [35]:

$$h_k(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x, D = k | X \geq x)}{\Delta x}.$$

In the case of independent competing events, the cause-specific and marginal hazard rates coincide [22, Ch. 3.5].

The cumulative incidence function is a common measure to quantify the difference in risk between the competing events. It describes the total probability of experiencing a specific event $k \in K$ up until time x:

$$F_k(x) = \mathbb{P}(X \leq x, D = k) = \int_0^x h_k(u) e^{\sum_{i \in K} \int_0^x h_i(v) dv} du.$$

The cumulative incidence is directly estimable from the data, since it uses only the cause-specific hazard functions and makes no assumptions on the joint distribution

of the failure times [22, Ch. 3.5]. At each moment in time, the cumulative incidence for each $k \in K$ cannot exceed 1, nor can the sum of cumulative incidences over all $k \in K$.

Instead of handling competing events as such in the model, we can also consider all events that are not the primary event of interest as censored observations. However, treating competing risks as censored observations violates the independence assumption between the censoring and event time distributions. As such, we will find a bias in models that assume these two distributions to be independent, such as the Kaplan-Meier estimator. Models that make no such assumptions, such as the Cox regression model, are still appropriate [35].

2.8 Bias

Lastly, like in all areas of research, bias is an important topic to consider in survival analysis. Apart from the common biases, such as selection, information and recall bias, several biases exist that are unique to survival analysis.

To start with, immortal time bias describes the situation where groups are made based on information that was unknown at baseline, making individuals in one group immortal until the moment they experience the event on which the groups were made [23, Ch. 21.1.1]. A famous example that demonstrates this bias is the study that concluded Oscar winners lived longer than those who had not won an Oscar [38]. Here, winning an Oscar or not was information that was unknown at baseline (i.e. birth), making those in the Oscar-winning group immortal until the moment of winning. A way to resolve this type of bias is to work with time-dependent covariates, where covariates are not set from the baseline measurement, but can change over time.

Secondly, two types of biases are distinguished which may contribute to an overestimation of the survival. The first of these biases is the lead time bias, which describes the instance where due to earlier detection of the baseline state the survival is overestimated. For example, since the introduction of screening for colorectal cancer, the time until death since diagnosis of colorectal cancer has increased. This, however, is not necessarily due to improved healthcare, but is rather due to inclusion of individuals at an earlier stage of the disease than was previously possible. As such, the survival seems to increase. Similarly, length time bias describes how slowly progressing states are more likely to be diagnosed, leading to an overestimation of the overall survival. Namely, since slowly progressing states have a higher probability of being sampled due to residing in this state for a longer time, they may be disproportionately represented in the sample, which will overestimate the survival [12].

3 Multi-state models

This chapter deals with the construction of multi-state models and their estimation.

A multi-state model is a generalization of the singular time-to-event situation, where intermediate stages of a process, or multiple outcomes, can be modelled simultaneously. This type of model may be interesting for the modelling of diseases with several stages, or to describe how patients may experience different outcomes after a shared diagnosis. The simplest multi-state model consists of three states, with three transitions and no cycles, see Figure 2. This model, also called the illness-death model, will be discussed in Section 3.4. Another simple model is the progressive model as seen in Figure 3, which allows only a singular option to exit a state in a one-way traffic fashion.

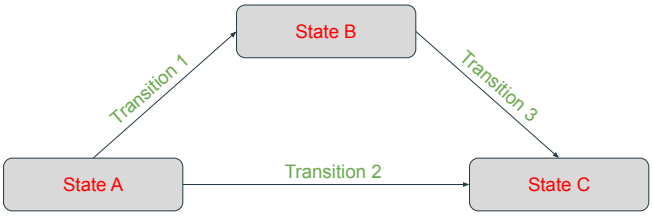


Figure 2: Three-state model.

Figure 3: Progressive multi-state model.

Multi-state models are stochastic processes, in which each outcome or stage in the process is called a *state*, as denoted in red in Figure 2. *Transitions* going from one state to another are denoted in green. States take values in the *state space* $\mathcal{S} = \{1, ..., s\}$ and can either be *starting*, *transient* or *absorbing*. Intuitively, starting states are states in which individuals start the process described in the model, transient states are states which individuals may enter and exit, and absorbing states are states which cannot be exited once entered. Using the transitions, we typically see that starting states only have transitions leaving the state, transient states have transitions going in and out of the state, and absorbing states only have transitions going into the state. In our three-state model of Figure 2, state A, B and C would be the starting, transient and absorbing state, respectively. The time at which individuals enter a state is called the *entering time*, the time at which individuals leave a state is called the *exit time*, and the time interval in which an individual has occupied a certain state is called the *sojourn time*.

The trajectory of individuals as they go through states is described as the *history*. Let $Q(x)$ denote the occupied state at time x for a given individual, then the history at time x for that individual is defined as

$$\mathscr{H}_x = \{Q(u) : x_0 \leq u \leq x\}.$$

With the generalization to a multiple-state setting, generalizations of the basic quantities described in Section 2 are also needed. More specifically, where the

(cumulative) hazard function was previously defined for a single event of interest, each transition is now an event in itself. As such, the (cumulative) hazard needs a transition-specific definition. Let X be the time at which state j is reached from state i, with $i, j \in \{1, ..., s\}$. Then, the hazard rate and cumulative hazard for a transition from state i to state j are defined as

$$h_{ij}(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x | X \geq x)}{\Delta x},$$

$$H_{ij}(x) = \int_0^x h_{ij}(u)du.$$

The hazard rate is now alternatively called the *transition intensity*. Similarly to the cumulative hazard for a transition, the *total hazard* to move out of a state can be estimated by summing over all hazards going out of the state [18]. The total hazard and total cumulative hazard to leave state i are then, respectively, given by

$$h_{i\bullet}(x) = \sum_{\mathcal{S} \ni j \neq i} h_{ij}(x),$$

$$H_{i\bullet}(x) = \int_0^x h_{i\bullet}(u)du.$$

The purpose of multi-state models are twofold. Firstly, the effect of covariates on the transition intensities can be estimated. Secondly, using the transition-specific covariate effects, predictions can be made on experiencing a transition for an individual with known covariates. This may be useful in, for example, personalized medicine approaches. Several methods for the estimation of covariate effects, as well as for prediction, exist, the most common of which will be discussed in the sections below. Alternatively, non-parametric approaches may be used in the estimation of the transition probabilities, in which no covariate information is used.

For the remainder of this chapter, we will assume event times to be measured continuously, independence between event and censoring times and right censoring to be the only censoring mechanism present in the data.

3.1 Markov assumption

Before moving on to the estimation, an important assumption - the Markov assumption - needs to be discussed. The Markov assumption, in the case of MSMs, says that making the transition to a future state is dependent only on the current state and the current time, but not on any previous states. In other words, any estimation depends only on the current state and time.

The implementation of the Markov assumption in MSMs comes in three flavours: the Markovian, semi-Markovian and non-Markovian models. As the name suggests, the former two use (adaptations) of the Markov assumption, while the latter is devoid of Markovianity. In the Markovian models, it is assumed that entering a new states warrants any information about all previous states unnecessary. In these types of models, the Markov assumption can be explicitly checked by for example using the entry time into each state as a covariate in the model that estimates

the (transition-specific) covariate effects on the transition intensities, or by using a stratified Commenges-Andersen score test of homogeneity [44]. Estimation of Markovian models use the so-called *clock forward* approach, in which the time starts at x_0 in the starting state and keeps continuously counting thereafter. Alternatively, in the *clock reset* method, time starts at x_0 in the starting state, after which the time is reset to zero each time a new state is entered. In doing so, any information about a previous state is essentially lost, imposing adherence to the Markov assumption in the model [35]. Thereby, models in which the Markovian assumption is violated, may use the clock reset method to be able to use a Markovian approach nonetheless.

Closely related to the Markovian models, are the semi-Markovian or Markov extension models. In these models, the sojourn time of the current state as well as the current state are considered in the analysis [18]. In the non-Markovian models, interest is in using the full history of an individual's previous state and entering/exit times in the estimation of the transition probabilities. Historically, since incorporation of the history into the estimation the transition probabilities complicates the information structure, estimation in Non-Markovian models relied on simulations. However, recently, more interest has been shown in the estimation of transition probabilities in non-Markovian settings. Further information on these types of estimations can be found in [3, 14, 30, 31, 43].

Tying in with the Markov assumption, a distinction is made between the time-homogeneous and the time-inhomogeneous settings. Here, time-homogeneity describes models in which the probability to make a certain transition is constant over time. Contrastingly, in a time-inhomogenous setting the transition probabilities change over time, making them time-dependent.

In the remainder of this chapter, we will be assuming a time-inhomogeneous Markovian setting.

3.2 Estimation of covariate effects

The first step in estimating the transition probabilities, is to estimate the contributions of the covariates of interest. Since non-parametric methods do not allow for incorporation of extra information, covariate effects can only be determined by using parametric or semi-parametric approaches. In this section, we focus on estimating covariate effects using the Cox proportional hazards regression model.

3.2.1 Likelihood

Before moving on to the likelihood based on the Cox model, we introduce the general likelihood for multi-state models. Let $i \in \{1, ..., m\} = \mathcal{I}$ denote the individuals and let n_i be the number of visited states for a given individual i. Then, we define $s_{ij} \in \mathcal{S}$ as the j^{th} state occupied by individual i, and $x_{ij} \geq 0$ as the time at which individual i entered the j^{th} state.

For each individual i, the contribution to the likelihood is determined by considering all experienced transitions, where each transition has a contribution to the likelihood of the following form:

$$h_{s_{ij},s_{i,j+1}}(x_{i,j+1})e^{-\left(H_{s_{ij}}(x_{i,j+1})-H_{s_{ij}}(x_{ij})\right)}, \tag{2}$$

where we see that the contribution to the likelihood consists of the probability to make the transition between state j and state $j+1$ at time $x_{i,j+1}$ (the hazard) and the probability to survive in state j from its entering time x_{ij} until the exit time $x_{i,j+1}$.

Since several transitions can be made from a state, the expression in (2) is repeated for all possible subsequent states $\ell \in \mathcal{S}$. The total contribution to the likelihood from a given state s_{ij} for individual i is then given by

$$\prod_{\mathcal{S} \ni \ell \leftarrow s_{ij}} h_{s_{ij},\ell}(x_{i,j+1})^{\mathbf{1}\{s_{i,j+1}=\ell\}} e^{-\left(H_{s_{ij}}(x_{i,j+1}) - H_{s_{ij}}(x_{ij})\right)}, \tag{3}$$

where the indicator function $\mathbf{1}\{s_{i,j+1} = \ell\}$ is used to make sure only transitions that were actually undergone by the individual contribute to the likelihood.

The likelihood of the full model can then be found by multiplying the contributions to the likelihood of all individuals i, from all states, where each of these contributions is of the form as seen in (3). The likelihood of the full model in the case of the final state being an absorbing state, is then equal to ([18, 34])

$$\begin{aligned} L = & \prod_{\mathcal{S} \ni \ell \leftarrow s_{i0}} h_{s_{i0},\ell}(x_{i1})^{\mathbf{1}\{s_{i1}=\ell\}} e^{-\left(H_{s_{i0},\bullet}(x_{i1}) - H_{s_{i0},\bullet}(x_{i0})\right)} \\ & \cdot \prod_{\mathcal{S} \ni \ell \leftarrow s_{i1}} h_{s_{i1},\ell}(x_{i2})^{\mathbf{1}\{s_{i2}=\ell\}} e^{-\left(H_{s_{i1},\bullet}(x_{i2}) - H_{s_{i1},\bullet}(x_{i1})\right)} \\ & \qquad \cdots \\ & \cdot \prod_{\mathcal{S} \ni \ell \leftarrow s_{i,n_i-1}} h_{s_{i,n_i-1},\ell}(x_{i,n_i})^{\mathbf{1}\{s_{i,n_i}=\ell\}} e^{-\left(H_{s_{i,n_i-1},\bullet}(x_{i,n_i}) - H_{s_{i,n_i-1},\bullet}(x_{i,n_i-1})\right)}. \end{aligned}$$

Should the final state not be an absorbing state, then an extra term is needed in the last product, which accounts for the right-censoring time c_i [34]. The likelihood is then as follows, with the extra term added in blue

$$\begin{aligned} L = & \prod_{\mathcal{S} \ni \ell \leftarrow s_{i0}} h_{s_{i0},\ell}(x_{i1})^{\mathbf{1}\{s_{i1}=\ell\}} e^{-\left(H_{s_{i0},\bullet}(x_{i1}) - H_{s_{i0},\bullet}(x_{i0})\right)} \\ & \cdot \prod_{\mathcal{S} \ni \ell \leftarrow s_{i1}} h_{s_{i1},\ell}(x_{i2})^{\mathbf{1}\{s_{i2}=\ell\}} e^{-\left(H_{s_{i1},\bullet}(x_{i2}) - H_{s_{i1},\bullet}(x_{i1})\right)} \\ & \qquad \cdots \\ & \cdot \prod_{\mathcal{S} \ni \ell \leftarrow s_{i,n_i-1}} h_{s_{i,n_i-1},\ell}(x_{i,n_i})^{\mathbf{1}\{s_{i,n_i}=\ell\}} e^{-\left(H_{s_{i,n_i-1},\bullet}(x_{i,n_i}) - H_{s_{i,n_i-1},\bullet}(x_{i,n_i-1})\right)} \\ & \cdot e^{-\left(H_{s_{i,n_i-1},\bullet}(c_i) - H_{s_{i,n_i-1},\bullet}(x_{i,n_i})\right)}. \end{aligned} \tag{4}$$

Optimization of this likelihood leads to the model that is most likely according to the data.

In the literature, many likelihoods have been derived for multi-state models that all differ slightly, depending on the situation at hand. For example, the number of

states, the censoring and truncation mechanisms, adherence to the Markov assumption and the type of estimator used for the hazard all influence how the likelihood may be simplified or rewritten. In the remainder of this subsection, we will focus on the likelihood of a Cox model.

As seen in Section 2.6, the Cox model given a covariate matrix Z is as follows

$$h(x|Z) = h_0(x)e^{Z^T\beta}.$$

The conditional probability that individual i undergoes the event at time x, denoted by x_i, can be expressed by ([7, Ch. 8]):

$$\begin{aligned}\mathbb{P}(x_i|\beta, Z) &= \frac{h_0(x)e^{Z_i^T\beta}}{\sum_{x_j \geq x_i} h_0(x)e^{Z_j^T\beta}} \\ &= \frac{e^{Z_i^T\beta}}{\sum_{x_j \geq x_i} e^{Z_j^T\beta}},\end{aligned}$$

where Z_i is the vector of covariate values for individual i, assuming the covariate values are time-fixed, and $x_j \geq x_i$ denotes the inclusion of all individuals j whose event time is larger than that of individual i - i.e. the set of individuals who are still under observation at a time just prior to x_i. Note that the second equality is due to the proportionality in the Cox model, where the baseline hazards are identical for each individual, causing them to be cancelled in the division. The partial likelihood then equals ([7, Ch. 8])

$$L(\beta|x, Z) = \prod_i^D \frac{e^{Z_i^T\beta}}{\sum_{x_j \geq x_i} e^{Z_j^T\beta}}, \tag{5}$$

where D is the number of events.

To estimate the regression coefficients $\hat{\beta}$ - which quantify the covariate effects - in this setting, the partial likelihood is maximized. There may be situations in which analytical solutions of this optimization are possible, but generally numerical optimization methods are used. In particular, the `coxph()`-function from the `survival` package in `R` uses the Newton-Raphson method to estimate the regression coefficients [36, 42].

A distinction can be made between overall covariate effects and transition-specific covariate effects through the design matrix Z. Namely, transition-specific covariates can be fit by measuring the effect of each covariate separately for all transitions. For example, if interest is in the two covariates gender and age, then these should be recorded separately for each transition. In the case of a model with three transitions, this means a total of 6 columns are needed to record gender and age. In this way, through addition of each separate column in the likelihood, the effect of the covariate on specific transitions can be determined, thereby determining the transition-specific covariate effects. This general model is equivalent to performing Cox regression on each transition separately.

As such, the Cox proportional hazards regression model can be used in a multi-state setting by using a design matrix in which all covariates are recorded separately

for each transition. Then, optimization of the partial likelihood shown in (5), with regards to the regression coefficients β, results in estimates for the transition-specific covariate effects.

3.3 Transition probabilities and their estimation

The second important aspect of multi-state models is their ability to estimate the probability to make a transition to a certain state at any given time. The estimation of these probabilities, the so-called transition probabilities, have been researched in a multitude of possible scenarios (and combinations thereof): time-homogeneous Markovian, time-inhomogeneous Markovian, semi-Markovian, non-Markovian, parametrically, non-parametrically, and so on, using different estimators, such as the Nelson-Aalen and Aalen-Johansen estimator [3; 8; 32; 33; 2, Ch. IV.4]. In the remainder of this section, we will be assuming a time-inhomogeneous Markovian setting.

Let $i, j \in \mathcal{S}$, and x, y any time after x_0 with $x \leq y$, where x_0 is the starting time. Then, a *transition probability* is defined as

$$\mathbb{P}_{ij}(x, y) = \mathbb{P}(Q(y) = j | Q(x) = i, \mathcal{H}_{x^-}).$$

Using the Markov property, this can be rewritten as

$$\mathbb{P}_{ij}(x, y) = \mathbb{P}(Q(y) = j | Q(x) = i). \tag{6}$$

Since a time-inhomogeneous setting was assumed, a simplification in the transition probability to start from zero cannot be made. For $x_0 \leq x \leq y$ and $i, j \in \mathcal{S}$ it holds that

$$\mathbb{P}_{ij}(x, y) \neq \mathbb{P}_{ij}(0, y - x).$$

Now, define three matrices to describe (i) the transition probabilities from time x to y, (ii) the transition intensities at time x, and (iii) the cumulative transition intensities at time x. These three matrices are respectively defined as

$$P(x, y) = \{\mathbb{P}_{ij}(x, y) : i, j \in \mathcal{S}, x_0 \leq x \leq y\},$$

$$h(x) = \{h_{ij}(x) : i, j \in \mathcal{S}, x_0 \leq x\},$$

and

$$H(x) = \{H_{ij}(x) : i, j \in \mathcal{S}, x_0 \leq x\}.$$

Define a fourth matrix,

$$\hat{H}(x) = \left\{\hat{H}_{ij}(x) : i, j \in \mathcal{S}, x_0 \leq x\right\},$$

which contains the estimated quantities of the cumulative transition intensities at time x. In theory, any estimator of the cumulative hazard could be used, depending on which is most suitable for the application at hand. Generally, the diagonal elements of $\hat{H}(x)$ are estimated as

$$\hat{H}_{ii}(x) = 1 - \sum_{\mathcal{S} \ni j \neq i} \hat{H}_{ij}(x).$$

3.3.1 Chapman-Kolmogorov

The transition probabilities as defined in (6) satisfy the Chapman-Kolmogorov equations [15, Theorem 12.13], which state that for $i, j \in \mathcal{S}$, $x_0 \leq x \leq y$, it holds that

$$\mathbb{P}_{ij}(x, y) = \sum_{k \in S} \mathbb{P}_{ik}(x, u)\mathbb{P}_{kj}(u, y). \tag{7}$$

From which it also follows that

$$P(x, y) = P(x, u)P(u, y). \tag{8}$$

Property (7) can be easily derived. To do so, let $x_0 \leq x \leq u \leq y$ and $i, j, k \in \mathcal{S}$, then

$$\begin{aligned}
\mathbb{P}_{ij}(x, y) &= \mathbb{P}(Q(x) = j | Q(s) = i) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}(Q(x) = j | Q(u) = k, Q(s) = i)\mathbb{P}(Q(u) = k | Q(s) = i) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}(Q(x) = j | Q(u) = k)\mathbb{P}(Q(u) = k | Q(s) = i) \\
&= \sum_{k \in \mathcal{S}} \mathbb{P}_{ik}(x, u)\mathbb{P}_{kj}(u, y).
\end{aligned}$$

Additionally, property (8) is now a direct consequence of (7) and basic linear algebra:

$$\begin{aligned}
P(x, y) &= \{\mathbb{P}_{ij}(x, y)\} \\
&= \left\{ \sum_{k \in S} \mathbb{P}_{ik}(x, u)\mathbb{P}_{kj}(u, y) \right\} \\
&= P(x, u)P(u, y).
\end{aligned}$$

3.3.2 Estimation

Due to the transition probabilities satisfying the Chapman-Kolmogorov equation, the Chapman-Kolmogorov *forward* equation − the infinitesimal counterpart of the Chapman-Kolmogorov equation − can be applied, which says that $\frac{\delta}{\delta t}P$ has a unique solution [18, 33]. This solution is given by

$$\frac{\delta}{\delta x} P(x, y) = P(x, y)h(x).$$

Using Volterra's integral equation [33, 34], we find the unique solution

$$P(x, y) = \prod_{(x,y]} (I + dH(u)).$$

Then, by using the estimated cumulative transition intensities, $P(x, y)$ can be estimated as follows

$$\hat{P}(x, y) = \prod_{x < u \leq y} (I + \Delta \hat{H}(u)).$$

In theory, any estimate of the $H_{ij}(u)$ could be used for the construction of $\hat{H}(u)$. But, the Aalen-Johansen estimator has been particularly defined, and arises when the $H_{ij}(u)$ are estimated using the Nelson-Aalen estimator shown in (1) [23, Ch. 8.3; 2, Ch. IV.4]. In our application, we will use the Cox estimates of the hazard, the derivation of which was discussed in Section 3.2.

The transition probabilities can be also transformed to the so-called state occupation probabilities, which describe the total probability to occupy a certain state at a given time. Visualisation of the state occupation probabilities over time are a popular representation of the transition probabilities.

3.4 Illness-death model

The illness-death model (Figure 4), also sometimes referred to as the disability model, is one of the simplest multi-state models. This model has three states: healthy, ill and deceased, which together a depict a simplified life-cycle in which individuals always start out healthy, may become diseased and always eventually die. It also assumes that sick individuals cannot become healthy again. Thereby, the model has three transitions, with 'Healthy' as the starting state, 'Illness' as a transient state and 'Death' as an absorbing state. For the illness-death model, it is possible to explicitly derive the transition probabilities.

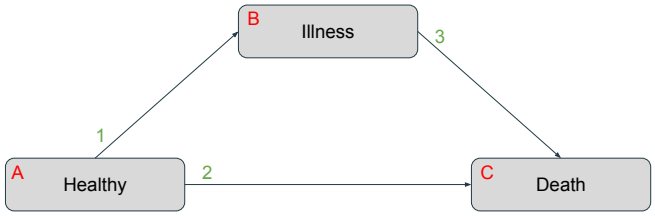


Figure 4: Illness-death model, with red and green indicating state and transition labels, respectively.

To express the transition probabilities at time u, we evaluate the probability to make the transition from state i to state j $(i, j \in \mathcal{S})$ before a given time x $(u \leq x)$. We will evaluate the transition probabilities for all possible combinations $i, j \in \mathcal{S}$. Let R be the random variable denoting the event of making the transition to state B, and X as the random variable denoting the event of making a transition to state C. Furthermore, we assume the Markov property to hold, meaning that the clock forward approach is implemented. The steps used in the calculations below have been previously described by Putter, Fiocco and Geskus (2007) [35]. Alternative approaches can be found in [16] and [45], showing analyses of the Aalen-Johansen estimator in a semi- and non-Markovian setting, and a non-parametric approach in a cross-sectional study sample, respectively.

We will estimate the transition probabilities from the easiest to the hardest to calculate. Therefore, we start in state C; since C is an absorbing state, leaving this state has probability 0. As such, the probability to remain in state C is given by

$$\mathbb{P}_{\mathrm{CC}}(u, x) = \mathbb{P}(Q(x) = \mathrm{C} | Q(u) = \mathrm{C}) = 1.$$

Moving on to the transition probabilities starting from state B. In state B, two options are possible: either staying in state B, or making the transition to state C. Since in this model it is assumed that only state A acts as a starting state, it is also necessary that the transition from A to B is made prior to being in state B. In order to incorporate this, we let r denote the entering time of state B and assume that the transition to state B from state A has been made. Then, starting with the transition between B and C, we know that to make this transition at time $v \in [u, x]$, it is necessary to survive in state B until time v after which the transition between the two states needs to occur. This can be expressed by the product

$$h_{\mathrm{BC,r}}(v)e^{-\int_u^v h_{\mathrm{BC,r}}(w)}. \tag{9}$$

Furthermore, we know that

$$\begin{aligned} S_{\mathrm{B,r}}(v) &= e^{-\int_0^v h_{\mathrm{BC,r}}(w)dw} \\ &= e^{-(\int_0^u h_{\mathrm{BC,r}}(w)dw + \int_u^v h_{\mathrm{BC,r}}(w)dw)} \\ &= e^{-\int_0^u h_{\mathrm{BC,r}}(w)dw} e^{-\int_u^v h_{\mathrm{BC,r}}(w)dw)} \\ &= S_{\mathrm{B,r}}(u)e^{-\int_u^v h_{\mathrm{BC,r}}(w)dw}, \end{aligned}$$

which implies that

$$e^{-\int_u^v h_{\mathrm{BC,r}}(w)dw} = \frac{S_{\mathrm{B,r}}(v)}{S_{\mathrm{B,r}}(u)}. \tag{10}$$

Additionally, we will use that the expression

$$S_{\mathrm{B,r}}(u) - S_{\mathrm{B,r}}(x) = \mathbb{P}(u \leq X \leq x), \tag{11}$$

expresses the probability that the transition between B and C will occur between time u and x. Since this is exactly what is portrayed by $\int_u^x h_{\mathrm{BC,r}}(v)S_{\mathrm{B,r}}(v)dv$, this can be substituted into the expression. Integrating over all possible values for v then gives the total probability to make the transition between state B and C:

$$\begin{aligned} \mathbb{P}_{\mathrm{BC,r}}(u, x) &= \int_u^x \mathbb{P}(X \leq v | X > u, R = r \leq u)\, dv \\ &\overset{(9)}{=} \int_u^x h_{\mathrm{BC,r}}(v)e^{-\int_u^v h_{\mathrm{BC,r}}(w)dw}dv \\ &\overset{(10)}{=} \frac{\int_u^x h_{\mathrm{BC,r}}(v)S_{\mathrm{B,r}}(v)dv}{S_{\mathrm{B,r}}(u)} \\ &\overset{(11)}{=} \frac{S_{\mathrm{B,r}}(u) - S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)} \\ &= 1 - \frac{S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)}. \end{aligned}$$

From $\mathbb{P}_{\mathrm{BC,r}}(u,x)$, we can also easily derive the probability to stay in state B:

$$\begin{aligned}\mathbb{P}_{\mathrm{BB,r}}(u,x) &= \mathbb{P}(X > x | X > u, R = r \leq u)\\ &= 1 - \mathbb{P}_{\mathrm{BC,r}}(u,x)\\ &= 1 - (1 - \frac{S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)})\\ &= \frac{S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)}.\end{aligned}$$

Lastly, we need to calculate the transition probabilities from the starting state A. Three transitions are possible from state A: staying in state A, leaving to state B, or leaving to state C. First, we will first discuss the transition from state A to B, then the transition from state A to C, and lastly remaining in state A. To make the transition from A to B, at some time $r \in [u,x]$, three things need to happen: survival in state A from time u until time r, making the transition between A and B at time r, and staying in state B until time x. These three steps can be represented by the survival in state A, the hazard to make the transition between A and B, and the probability to stay in state B, respectively. This gives the expression

$$e^{-\int_u^r h_{\mathrm{AB}}(v)dv} h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BB,r}}(r,x), \tag{12}$$

where $\mathbb{P}_{\mathrm{BB,r}}(r,x)$ is the 'transition' probability to stay in state B. Moreover, we again use that

$$e^{-\int_u^r h_{\mathrm{AB}}(v)dv} = \frac{S_{\mathrm{A}}(r)}{S_{\mathrm{A}}(u)}, \tag{13}$$

as derived previously for Equation (10). Integrating over all possible values of r then gives $\mathbb{P}_{\mathrm{AB}}(u,x)$:

$$\begin{aligned}\mathbb{P}_{\mathrm{AB}}(u,x) &= \int_u^x \mathbb{P}(R \leq r < X | R > u, X > u)\, dr\\ &\overset{(12)}{=} \int_u^x e^{-\int_u^r h_{\mathrm{AB}}(v)dv} h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BB,r}}(r,x)\, dr\\ &\overset{(13)}{=} \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BB,r}}(r,x)\, dr}{S_{\mathrm{A}}(u)}.\end{aligned}$$

The second possibility after starting in state A, is to make the transition to state C. This can be done in two ways, namely (I) directly from state A to state C, or (II) through state B. As such, to calculate $\mathbb{P}_{\mathrm{AC}}(u,x) = (I) + (II)$, these two alternatives need to be examined separately.

Starting with transition (I), which is similar to the transition probability between state A and B, with the exception that it is not necessary to include a term that describes the probability to remain in state C after the transition, due to C being an absorbing state. The other quantities can be similarly derived as before, based on the knowledge that to make the transition at time $v \in [u,x]$, it is necessary to survive in state A until time v after which the transition between the two states needs to occur. This can be expressed by

$$h_{\mathrm{AB}}(v) e^{-\int_u^v h_{\mathrm{AC}}(w)dw}. \tag{14}$$

Furthermore, in the simplification of the integral, we will also again use that

$$e^{-\int_u^v h_{\mathrm{AC}}(w)dw} = \frac{S_{\mathrm{A}}(v)}{S_{\mathrm{A}}(u)}. \tag{15}$$

Integrating over all possible values for v then gives the total probability to make the direct transition between state A and C:

$$\begin{aligned}(I) &= \int_u^x \mathbb{P}(X < v, X < R | X > u, R > u)\, dv \\ &\overset{(14)}{=} \int_u^x h_{\mathrm{AC}}(v) e^{-\int_u^v h_{\mathrm{AC}}(w)dw}\, dv \\ &\overset{(15)}{=} \frac{\int_u^x h_{\mathrm{AC}}(v) S_{\mathrm{A}}(v)\, dv}{S_{\mathrm{A}}(u)}.\end{aligned}$$

For the calculation of the probability to enter state C through state B, we will use the Chapman-Kolmogorov (CK) property, after which previously-derived transition probabilities can be inserted. This gives

$$\begin{aligned}(II) &= \int_u^x \mathbb{P}(R < X \leq r | R > u, X > u)\, dr \\ &\overset{(CK)}{=} \int_u^x \mathbb{P}_{\mathrm{AB}}(u, r) \mathbb{P}_{\mathrm{BC,r}}(r, x)\, dr \\ &= \int_u^x e^{-\int_u^r h_{\mathrm{AB}}(w)dw} h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r, x)\, dr \\ &\overset{(13)}{=} \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r, x)\, dr}{S_{\mathrm{A}}(u)}.\end{aligned}$$

The probability to enter state C from state A then equals

$$\begin{aligned}\mathbb{P}_{\mathrm{AC}}(u, x) &= (I) + (II) \\ &= \frac{\int_u^x h_{\mathrm{AC}}(r) S_{\mathrm{A}}(r)\, dr}{S_{\mathrm{A}}(u)} + \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r, x)\, dr}{S_{\mathrm{A}}(u)} \\ &= \frac{\int_u^x S_{\mathrm{A}}(r)(h_{\mathrm{AC}}(r) + h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r, x))\, dr}{S_{\mathrm{A}}(u)}.\end{aligned}$$

The last possibility when $Q(u) = \mathrm{A}$, is to stay in state A until time x. Since $\mathbb{P}_{\mathrm{AA}}(u, x) + \mathbb{P}_{\mathrm{AB}}(u, x) + \mathbb{P}_{\mathrm{AC}}(u, x) = 1$, we can find the probability to stay in state A by taking the complement of the probabilities we found before:

$$
\begin{aligned}
\mathbb{P}_{\mathrm{AA}}(u,x) &= 1 - (\mathbb{P}_{\mathrm{AB}}(u,x) + \mathbb{P}_{\mathrm{AC}}(u,x)) \\
&= 1 - (\mathbb{P}_{\mathrm{AB}}(u,x) + (I) + (II)) \\
&= 1 - \left(\frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BB,r}}(r,x)\, dr}{S_{\mathrm{A}}(u)} + \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r,x)\, dr}{S_{\mathrm{A}}(u)} + (I) \right) \\
&= 1 - \left(\frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) (\mathbb{P}_{\mathrm{BB,r}}(r,x) + \mathbb{P}_{\mathrm{BC,r}}(r,x))\, dr}{S_{\mathrm{A}}(u)} + (I) \right) \\
&= 1 - \left(\frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r)\, dr}{S_{\mathrm{A}}(u)} + \frac{\int_u^x h_{\mathrm{AC}}(r) S_{\mathrm{A}}(r) dr}{S_{\mathrm{A}}(u)} \right) \\
&= 1 - \frac{\int_u^x S_{\mathrm{A}}(r)(h_{\mathrm{AB}}(r) + h_{\mathrm{AC}}(r)) dr}{S_{\mathrm{A}}(u)} \\
&= 1 - \frac{S_{\mathrm{A}}(u) - S_{\mathrm{A}}(x)}{S_{\mathrm{A}}(u)} \\
&= \frac{S_{\mathrm{A}}(x)}{S_{\mathrm{A}}(u)}.
\end{aligned}
$$

We have now explicitly derived all analytic expression of the transition probabilities for the illness-death model, and they are given by:

$$
\begin{aligned}
\mathbb{P}_{\mathrm{AA}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{A} | Q(u) = \mathrm{A}) = \frac{S_{\mathrm{A}}(x)}{S_{\mathrm{A}}(u)}, \\
\mathbb{P}_{\mathrm{AB}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{B} | Q(u) = \mathrm{A}) = \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BB,r}}(r,x)\, dr}{S_{\mathrm{A}}(u)}, \\
\mathbb{P}_{\mathrm{AC}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{C} | Q(u) = \mathrm{A}) = \frac{\int_u^x h_{\mathrm{AC}}(v) S_{\mathrm{A}}(v) dv}{S_{\mathrm{A}}(u)} + \frac{\int_u^x S_{\mathrm{A}}(r) h_{\mathrm{AB}}(r) \mathbb{P}_{\mathrm{BC,r}}(r,x)\, dr}{S_{\mathrm{A}}(u)}, \\
\mathbb{P}_{\mathrm{BB,r}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{B} | Q(u) = Q(r) = \mathrm{B}, r \leq u) = \frac{S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)}, \\
\mathbb{P}_{\mathrm{BC,r}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{C} | Q(u) = Q(r) = \mathrm{B}, r \leq u) = 1 - \frac{S_{\mathrm{B,r}}(x)}{S_{\mathrm{B,r}}(u)}, \\
\mathbb{P}_{\mathrm{CC}}(u,x) &= \mathbb{P}(Q(x) = \mathrm{C} | Q(u) = \mathrm{C}) = 1.
\end{aligned}
$$

From the derivations it is also immediately apparent that the introduction of more states and/or more transition quickly complicates calculations. As such, analytical expressions for more complex (Markovian) MSMs are rare, and have little use in practice, especially in the abundance of good software than can reliably estimate these transition probabilities. In non-Markovian settings, bootstrap methods may be used for the estimation of the transition probabilities [14].

3.5 Extensions

As discussed before, the type of multi-state models and their estimation as described in this chapter are only a subset of all the possible flavours of this type of model.

Multiple other types of MSMs exist, such as the Bayesian multi-state models [24], and semi-Markovian [4] and non-Markovian models [30, 33, 43]. Furthermore, the hidden Markov models are a class of models that deal with sequence structures in a multi-state model fashion [13, 37]. Alternatively, while not as extensively discussed, the drifting Markov processes were defined to impose a smoothness on the time-inhomogeneous Markov models to deal with homogeneity in sequence structures [5, 46].

Moreover, several estimation methods for the estimation of the transition probabilities have been developed for a myriad of different multi-state models. The estimation of the transition probabilities as described in Section 3.3 is a general method that works for many Markovian models, in which the estimation of the transition intensities can be done as applicable to the situation. Other estimation methods may be useful for specific situations or other types of MSMs. For example, parametric estimation may be done using the Weibull distribution [28]. Additionally, a semi-parametric approach could be taken using splines with a penalized maximum likelihood to estimate the transition intensities [29], as well as a non- and semi-parametric landmark estimation approach [17].

4 Application to medical data: a data illustration

In this chapter, we will illustrate an application of multi-state models using a real-life data example.

This data illustration focuses on modelling the progression of Alzheimer's Disease (AD), using transition-specific covariates. AD is a progressive disease characterized by cognitive decline, episodic memory loss, and unpredictable behaviour [26]. Three phases are recognized in the progression to AD: cognitively normal (CN), mild cognitive impairment (MCI) and Alzheimer's Disease, respectively. This progression describes going from showing no cognitive decline at all (i.e. CN), to some cognitive decline (i.e. MCI), and lastly to the final stage dementia. Definitions for each mental state can be found in Appendix A. This application focuses on modelling the transitions between these three phases, each given its own state in the model. A fourth state, death, has furthermore been added, see Figure 5. Here, cognitively normal was considered as the starting state, MCI and dementia as the transient states, and death as the absorbing state.

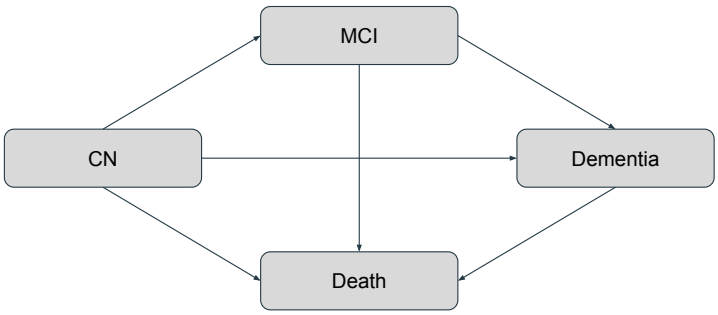


Figure 5: Four-state model showing the progression to dementia (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).

This type of model, which uses a MSM to model the progression to dementia, has been previously implemented in several studies, albeit with slightly different state spaces. In these studies, the focus was mostly on the estimation of (transition-specific) covariate effects, in which analytical derivations were done and simulation studies were performed. Namely, Salazar et al. (2007) and Yu et al. (2009) modeled the progression to dementia using a five-state model with CN and two types of cognitive impairment as the three transient states, where CN was additionally the starting state, and dementia and death as the two competing absorbing states [39, 49]. In their models a polytomous logistic regression model with shared random effects was used to estimate the effect of variables on the transitions. Kryscio et al. (2006) had previously similarly used polytomous logistic regression models to model the covariate effects, using the same multi-state model [25]. Wei, Xu and Kryscio (2014) built further on these ideas and researched how the parametric Weibull and semi-parametric Cox model with shared random effects performed on the five-stage multi-state model defined by Salazar et al. [48]. Wang et al. (2023) proposed a three-step estimation method to use fixed and random effects to estimate the covariate effects on the transition probabilities, effectively using a generalized linear mixed model approach [47]. Commenges and Joly (2004) created a model to evaluate the

interaction between dementia and institutionalization using a penalized likelihood approach [6]. This approach is elaborated further in the study described by Joly et al. (2009) [20]. Each of these studies focused on the estimation of the (transition-specific) covariate effects, and showed results with simulation studies. Most of these studies additionally showed an application using actual data.

With this data, we will similarly show an application of Markovian, time-inhomogeneous models in practice. However, instead of the aforementioned generalized linear mixed model approaches, the transition-specific covariate effects will be estimated using a Cox proportional hazards regression model. The transition-structure in our application also differs from those described before. Thereafter, the covariate effects will be applied to the estimation of the transition probabilities.

4.1 Data description and preparation

Data was obtained from the Alzheimer's Disease Neuroimaging Initiative, an ongoing, US-based, multi-center, longitudinal study:

> *Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) [19].*

ADNI aims at improving uniform data collection and data sharing in the field of Alzheimer's research. The multi-stage initiative, which commenced in 2004, is currently in it's third stage. For this thesis, we will use the aggregated data from all phases of the project. The three stages used the same inclusion criteria and can be found in Appendix B.

The data was obtained from the `adnimerge` dataset in the `ADNIMERGE` package in `R` [41]; death records were added from the `treatdis` dataset. Death records were identified by searching for the terms 'death', 'die' and 'expire' in the free text space of the `treatdis` dataset. After selection of individuals according to this procedure, these individuals were inspected manually to make sure their categorization into the death state was correct. Individuals for whom the date of death was obtained, were then added to the data acquired from `adnimerge`.

The `adnimerge` dataset consisted of 16302 observations in 2428 individuals. A total of 62 variables were recorded, of which 1 outcome, 2 identifying, 6 administrative, 6 non-medical and 47 medical variables. Categorization and further explanation on these variables can be found in Appendix C. The outcome variable consisted of the diagnosis of mental ability: cognitively normal, mild cognitive impairment or dementia. Note that no clear distinction was made between dementia and Alzheimer's Disease in this dataset. As such, statements will not be made on Alzheimer's Disease specifically, but rather on dementia as a whole.

Since interest is specifically in the mental state development over time, only individuals with at least 2 measurements have been included. Similarly, individuals without a baseline diagnosis - i.e. unknown starting state - were removed from analysis. Moreover, patients with improving mental states were removed from the data, due to such transitions being medically improbable. These transitions consisted of the transitions from dementia to either MCI or CN and from MCI to CN. Out of 2428 individuals, 1954 were eligible for analysis, see Figure 6. Patients were then categorized based on baseline diagnosis, leading to groups of 707, 880 and 367 individuals for the starting states CN, MCI and dementia, respectively.

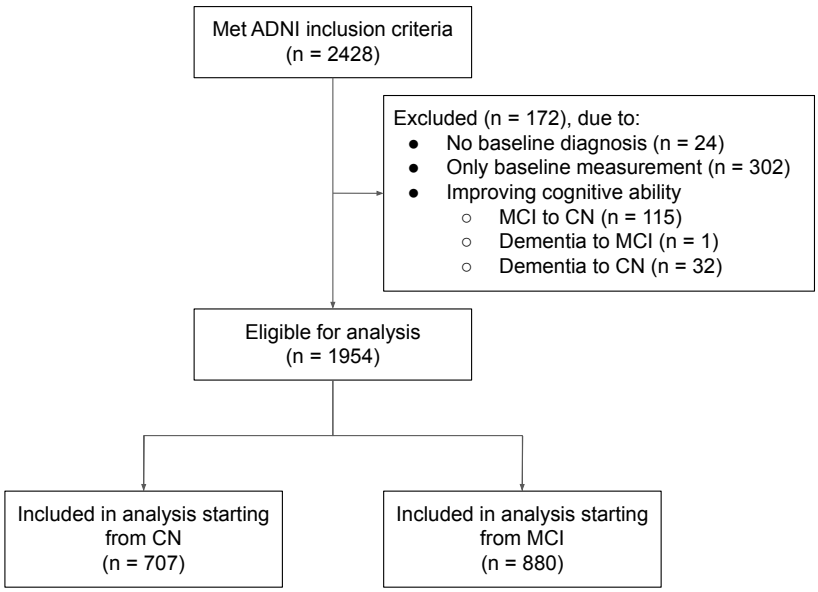


Figure 6: Consort diagram showing which ADNI data was included in analysis (ADNI = Alzheimer's Disease Neuroimaging Initiative, CN = Cognitively Normal, MCI = Mild Cognitive Impairment).

After exclusion of 474 individuals, 14256 observations remained, of which 4296 had no recorded diagnosis. This number was lowered to 702 by considering all consecutive *NA*-values of the same patient as a single *NA*-value. Similarly, CN, MCI and dementia were recorded 3380, 4280 and 2300, respectively. These counts were then reduced to 707, 999 and 750 respectively, after considering only the first moment of diagnosis for each state, per individual. Death was recorded 133 times.

To solve the missing state problem caused by the 702 missing diagnoses, it was firstly assumed that individuals with missing states between CN and MCI, or MCI and dementia, were in an intermediate state MCI−, or MCI+, respectively. Individuals with missing states after CN-diagnosis that had no further diagnoses, were censored after the last recorded CN-measurement. Similarly, missing states after dementia-diagnosis were interpreted as dementia. Ultimately, since entering the MCI− state was defined in a way that makes it dependent on its departure - namely, the MCI− can only be entered from the CN state if thereafter the MCI state was observed -, this extra state gives little information; leaving this state has probability 1. Thereby, it was decided to remove this state from the model and censor these patients after their last CN observation, leaving the final model to be a five-state

model with states CN, MCI, MCI+, dementia and death. This final model can be found in Figure 7.

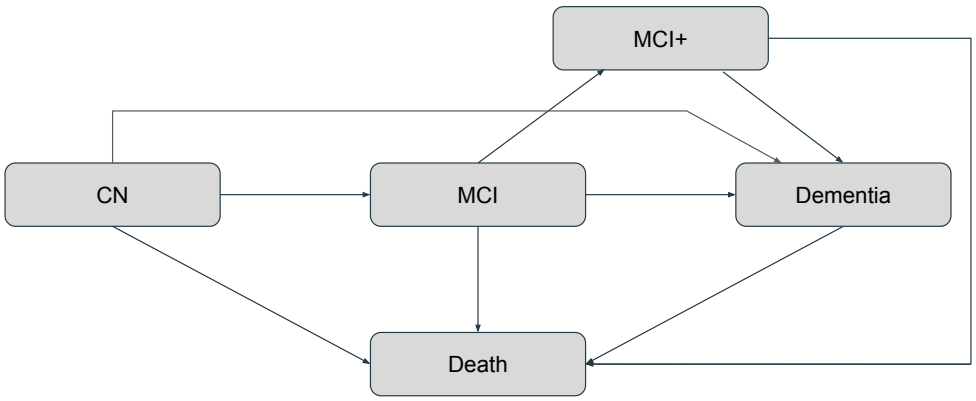


Figure 7: Multi-state model showing the progression to dementia (CN = cognitively normal, MCI = mild cognitive impairment).

4.2 Variable selection

As aforementioned, 6 non-medical and 47 medical variables were recorded. For this project, only variables that were measured at the first moment of each diagnosis were considered - i.e. measurements made at the entering times. Due to high percentages of missing values in the medical covariates, only covariates with $< 33\%$ missing were considered. See Appendix D for the percentage missing in all medical variables. In total, 22 of the medical variables had a missingness rate lower than 33% and were therefore eligible for analysis. Unlike the medical covariates, little missingness was observed for the non-medical variables. Namely, 3 (0.15%) missing values were recorded for the variable age and none for gender, education, ethnicity, race and marital status. The little missingness in these variables can be explained by the fact that these variables were time-fixed and are likely to have been recorded at baseline. Ethnicity and marital status were not considered in the analysis due to overlap with another variable (race) and non-relevance to the outcome variable, respectively. Summarizing statistics of the 22 medical and 4 non-medical variables as measured at baseline can be found in Table 1. This table shows many noticeable differences within variables between states on a sliding scale, with the MCI state showing intermediate values for the continuous variables.

Since many of the medical variables were of similar nature, co-linearity was expected to occur. Therefore, variables that showed strong correlations with other variables were removed. Variables that had an absolute correlation coefficient of more than 0.65 were inspected, after which the variable with the least missing values and the lowest correlation with other variables was kept. The correlation plot can be found in Figure 8. Out of the 22 medical variables that remained, 12 were removed due to co-linearity. Thereby, 10 medical and 4 non-medical covariates were considered to be appropriate for analysis.

Table 1: Summarizing statistics at baseline, per mental state.

	CN (n = 707)	**MCI** (n = 880)	**Dementia** (n = 367)
Gender = male (%)	351 (44.6)	526 (59.8)	207 (56.4)
Age in years	73.04 (6.22)	73.33 (7.38)	74.61 (7.88)
Education in years	16.51 (2.57)	15.93 (2.84)	15.22 (2.89)
Race (%)			
American Indian/Alaskan	1 (0.1)	2 (0.2)	0 (0.0)
Asian	12 (1.7)	15 (1.7)	8 (2.2)
Hawaiian/Other	0 (0.0)	2 (0.2)	0 (0.0)
Black	42 (5.9)	30 (3.4)	17 (4.6)
White	639 (90.4)	821 (93.3)	338 (92.1)
More than one	12 (1.7)	6 (0.7)	4 (1.1)
Unknown	1 (0.1)	4 (0.5)	0 (0.0)
MMSE	29.11 (1.13)	27.49 (1.85)	23.22 (2.15)
RAVLT			
Immediate	45.71 (9.92)	33.61 (10.23)	22.77 (7.21)
Learning	6.10 (2.33)	3.98 (2.54)	1.84 (1.81)
Forgetting	3.70 (2.81)	4.66 (2.56)	4.52 (1.82)
Percentage forgetting	34.22 (27.63)	61.78 (34.38)	89.67 (21.19)
TRABSCOR	82.04 (42.59)	120.17 (67.05)	195.26 (87.48)
FAQ	0.20 (0.85)	3.42 (4.25)	13.31 (6.89)
CDRSB	0.04 (0.14)	1.56 (0.91)	4.41 (1.69)
LDELTOTAL	13.20 (3.36)	5.60 (3.48)	1.43 (1.98)
ADAS11	5.66 (2.89)	10.55 (4.51)	19.49 (6.58)
ADAS13	8.77 (4.30)	17.04 (6.69)	29.85 (7.93)
ADASQ4	2.71 (1.81)	5.69 (2.47)	8.63 (1.50)
mPACC			
digit	0.05 (2.73)	-6.54 (3.90)	-14.94 (3.40)
trails B	0.05 (2.59)	-6.13 (3.78)	–14.18 (3.16)
APOE4 (%)			
0	485 (69.7)	411 (48.9)	111 (31.2)
1	193 (27.7)	332 (39.2)	171 (48.0)
2	18 (2.6)	104 (12.3)	74 (20.8)
Hippocampus $\cdot 10^2$	74.81 (8.91)	67.54 (11.13)	57.47 (10.04)
Ventricles $\cdot 10^3$	33.26 (18.31)	41.34 (22.38)	50.72 (24.66)
Fusiform $\cdot 10^3$	18.27 (2.47)	17.53 (2.74)	15.67 (2.66)
ICV $\cdot 10^5$	14.96 (1.59)	15.37 (1.63)	15.36 (2.29)
MidTemp $\cdot 10^3$	20.70 (2.76)	19.62 (2.98)	17.42 (3.02)
WholeBrain $\cdot 10^5$	10.42 (1.06)	10.29 (1.09)	9.79 (1.15)
Entorhinal $\cdot 10^2$	39.60 (6.75)	35.33 (7.83)	28.65 (7.39)

Statistics are presented as mean (standard deviation), unless indicated differently.
Abbreviations: CN = Cognitively Normal; MCI = Mild Cognitive Impairment; MSSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time; FAQ = Functional Assessment Questionnaire; CDRSB = Clinical Dementia Rating – Sum of Boxes; LDELTOTAL = Delayed Total Recall; ADAS11/13/Q4 = Alzheimer's Disease Assessment Scale 11 items/13 items/task 4; mPACCdigit/trailsB = Modified Preclinical Alzheimer Cognitive Composite with digit/trails B; APOE = Apolipoprotein E; ICV = Intra-Cranial Volume.

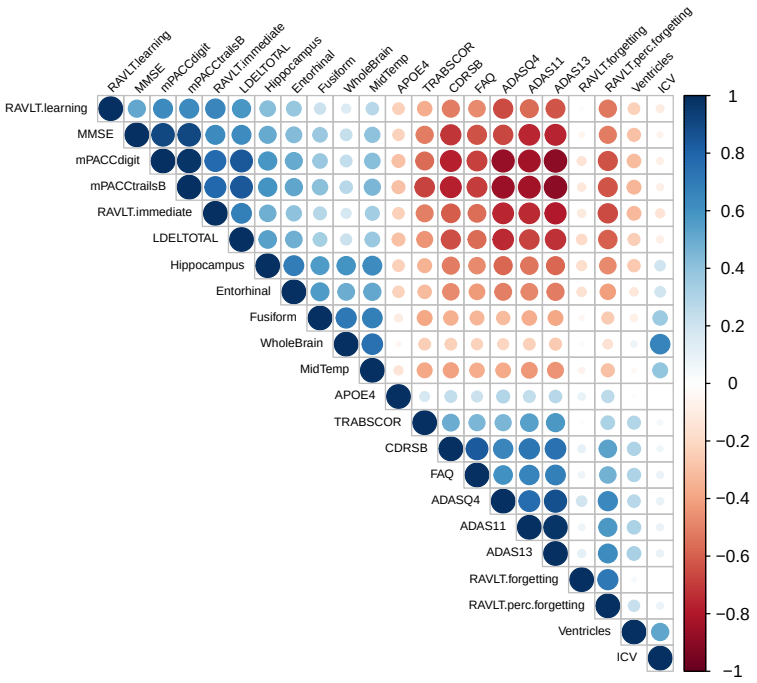


Figure 8: Correlation plot with the correlation coefficients between variables.

After an attempt to fit the first model using (a subset of) the above-mentioned 14 variables, it was decided that all numerical variables were to be categorized to deal with convergence problems. To do so, each variable was split into a lower and upper quantile, the former containing the lowest 50% of observations, the latter containing the highest 50% of observations (see Table 2).

Table 2: Distribution of the numerical variables into categorical variables, with both quantiles containing 50% of observations.

Variable	**Lower quantile**	**Upper quantile**
MMSE	$[7, 28]$	$(28, 30]$
RAVLT		
Immediate	$[1, 32]$	$(32, 73]$
Forgetting	$[-23, 4]$	$(4, 15]$
TRABSCOR	$[0, 98]$	$(98, 996]$
FAQ	$[0, 2]$	$(2, 30]$
Ventricles	$[5.65 \cdot 10^3, 3.73 \cdot 10^4]$	$(3.73 \cdot 10^4, 1.58 \cdot 10^5]$
Hippocampus	$[3.09 \cdot 10^3, 6.77 \cdot 10^3]$	$(6.77 \cdot 10^3, 1.08 \cdot 10^4]$
Fusiform	$[8.99 \cdot 10^3, 1.72 \cdot 10^4]$	$(1.72 \cdot 10^4, 3 \cdot 10^4]$
ICV	$[8.82 \cdot 10^5, 1.51 \cdot 10^6]$	$(1.51 \cdot 10^6, 3.32 \cdot 10^6]$

Abbreviations: MMSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time; FAQ = Functional Assessment Questionnaire; ICV = Intra-Cranial Volume.

4.3 Model fitting

Due to individuals entering the study in three possible starting states, an approach was needed which could take this structure into account. Separate multi-state models were fitted, with each model considering only those individuals that started in one specific state. Furthermore, since starting in the dementia state simplifies to a simple Cox regression problem, only the models starting from the CN and MCI state were fitted, because interest was specifically in inspecting the behaviour of multi-state models in this setting. Moreover, since individuals could have multiple observations within the time span of a single state, an attempt was made to include time-varying variables into the multi-state model. However, due to limitations in time and software, this was deemed not feasible. Thereby, the final analyses consisted of separate multi-state models starting from the CN and the MCI state at baseline, fitted both on the study and age scale; in total, four models were fitted. For all models the clock forward approach was used. The number of missing data at baseline for the used variables can be found in Table 3. Model fitting was done using the `mstate`-package in `R` (version 4.1.0, R studio version 2022.07.1 [9, 36, 40]), according to the tutorial as written by de Wreede, Fiocco and Putter (2011) [10]. The code written for the analysis can be found on this GitHub repository. A confidence level of $\alpha = 0.05$ was used.

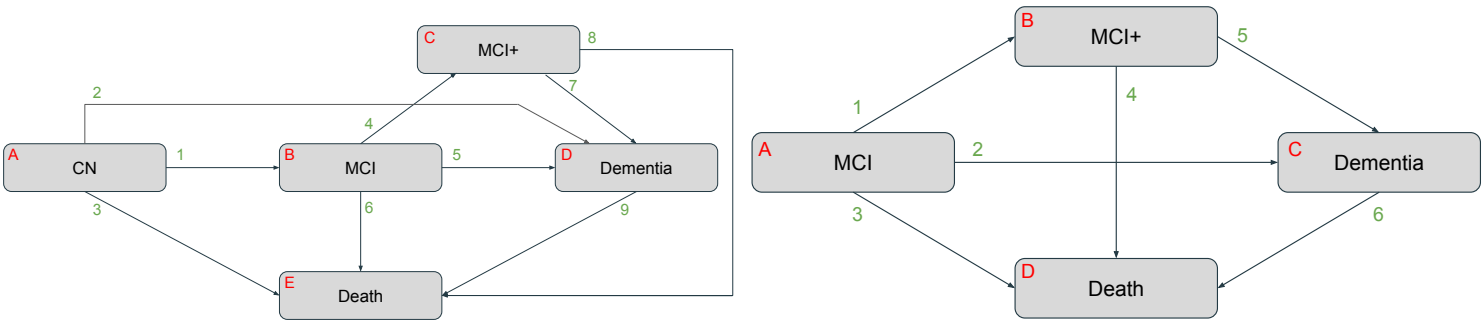
Table 3: Missingness for all included variables, per baseline category.

	CN (n = 707)	**MCI** (n = 880)
Gender	0 (0)	0 (0)
Age in years	0 (0)	3 (0.3)
MMSE	0 (0)	0 (0)
RAVLT		
Immediate	2 (0.3)	0 (0)
Forgetting	2 (0.3)	1 (0.1)
TRABSCOR	3 (0.4)	12 (1.4)
FAQ	0 (0)	12 (1.4)
APOE4	11 (1.6)	33 (3.8)
Ventricles	41 (5.8)	43 (4.9)
Hippocampus	70 (9.9)	135 (15.3)
Fusiform	75 (10.6)	138 (15.7)
ICV	19 (2.7)	22 (2.5)

Statistics are presented as number missing (%).
Abbreviations: MMSE = Mini-Mental State Exam;
RAVLT = Rey Auditory Verbal Learning Test;
TRABSCOR = Trail Making Test Part B Time;
FAQ = Functional Assessment Questionnaire;
APOE = Apolipoprotein E
ICV = Intra-Cranial Volume.

All steps described below apply to the models starting from the CN state, as well as the models starting from the MCI state. To fit the models, the data was firstly transformed from wide to long format in order to determine all transition-specific contributions to the likelihood. This formatting change is described only here, rather

than during the data preparation, because the way the data was transformed to long format depended on the transition matrices, which are model-specific. Note that the model starting from CN at baseline and the model starting from MCI at baseline were fundamentally different. Namely, where the former is a five-state model with nine transition, the latter has only four states and six transitions. State and transition labels can be found in Figure 9(a) and Figure 9(b), respectively, with red indicating state labels and green indicating transition numbers.



(a) CN as starting state

(b) MCI as starting state

Figure 9: Multi-state model showing the progression to dementia from different starting states, with red and green indicating state and transition labels, respectively (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).

Because the transition-structure differed between the two types of models, also two separate transition matrices were needed for the transformation of the data from wide to long format. These transition matrices can be found in (16) and (17), for the CN and MCI model, respectively.

$$\begin{array}{c} \\ \text{CN} \\ \text{MCI} \\ \text{MCI+} \\ \text{Dem} \\ \text{Death} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{CN} & \text{MCI} & \text{MCI+} & \text{Dem} & \text{Death} \end{array} \\ \begin{pmatrix} - & 1 & - & 2 & 3 \\ - & - & 4 & 5 & 6 \\ - & - & - & 7 & 8 \\ - & - & - & - & 9 \\ - & - & - & - & - \end{pmatrix} \end{array} \tag{16}$$

$$\begin{array}{c} \\ \text{MCI} \\ \text{MCI+} \\ \text{Dem} \\ \text{Death} \end{array} \begin{array}{c} \begin{array}{cccc} \text{MCI} & \text{MCI+} & \text{Dem} & \text{Death} \end{array} \\ \begin{pmatrix} - & 1 & 2 & 3 \\ - & - & 4 & 5 \\ - & - & - & 6 \\ - & - & - & - \end{pmatrix} \end{array} \tag{17}$$

Using the transition matrices, the datasets was transformed from wide to long format using `msprep()`, after which the covariates were expanded to be able to use them as transition-specific covariates by using `expand.cov()`. Thereafter, transition-specific covariates were chosen for the models, where each transition was given

approximately one covariate per ten events. The number of covariates per transition was therefore decided based on the number of observed events, after which the Cox model was fitted using the selected covariates.

After fitting the model, violation of the proportional hazards assumption was investigated using the `cox.zph()`-function from the `survival`-package. Thereafter, compliance to the Markov assumption was investigated by adding the entering times for MCI, MCI+, dementia and death to the models starting from CN at baseline and the entering times for MCI+, dementia and death to the models starting from MCI at baseline. The entering times were treated as transition-specific.

The transition probabilities were estimated using the `probtrans()`-function from the `mstate`-package. Transition probabilities were visualized and analyzed using state occupation probabilities, which describe the total probability to occupy a certain state and how this develops over time. State occupation probabilities were compared between patients with identical covariates, with exception of the covariate of interest. Comparisons were made for different ages and all categories of gender, MMSE and APOE4 alleles.

With regards to the models fitted on the age scale, all steps as described above apply, with the mere exception that the age at baseline was added to the observed times. By doing so, a comparison is made between patients of the same age, rather than between patients with the same number of years since inclusion in the study. This furthermore implied that age was not to be included as a covariate in the model on the age scale, since it was explicitly used in the observed time. Because the transition-structure and the number of events is invariant under the change of time scale within the CN and MCI models, the same number of covariates per transition were allowed on both scales. Therefore, the same covariates were used for the study and age scales, with exception of the age variable. After model fitting, the Markov assumption, as well as violation of the proportional hazards assumption were investigated as previously described. Lastly, the transition probabilities were estimated.

Outcomes of the model starting from CN at baseline on the study and age scale, respectively, and the model starting from MCI at baseline on the study and age scale, respectively, will be discussed in Sections 4.3.1 to 4.3.4.

4.3.1 Model starting from the CN state (study scale)

For the multi-state model starting from the CN state, the number of events per transition can be found in Figure 10, with blue and orange indicating the number of individuals starting and ending in that state, respectively and black indicating the number of individuals that have undergone the transition.

Transitions 1, 3, 4, 5 and 7 could be fitted with 12, 2, 3, 2 and 1 covariate(s), respectively. For transitions 2, 6, 8 and 9, no covariates could be fitted due to the low number of events. Transition-specific covariates were chosen on the basis of theoretical knowledge, as well as trial and error with respect to convergence of the model: due to the low number of events per transition, the model showed convergence issues when combining specific covariates.

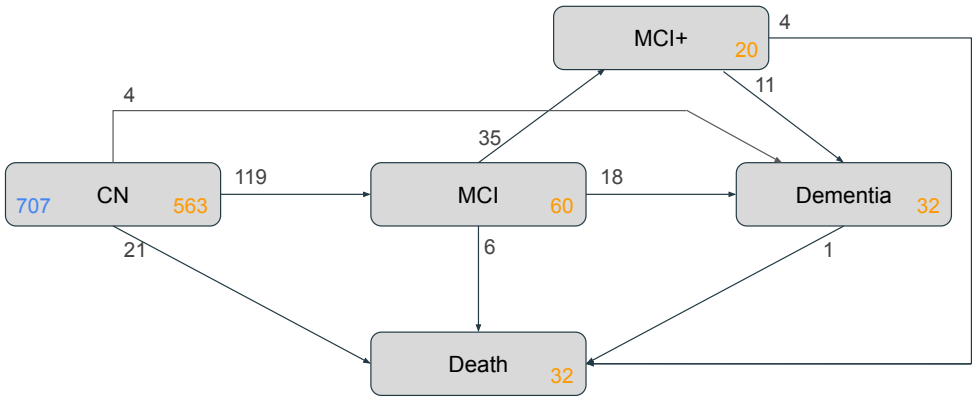


Figure 10: Multi-state model showing the progression to dementia starting from CN, with blue and orange indicating the number of people starting and ending in a state, and black indicating the number of people undering the transition (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).

After fitting the model, adherence to the Markov assumption was investigated by adding the entering times for MCI, MCI+, dementia and death to the model. Adding any subset of these entering times lead to non-convergence of the model. Therefore, the Markov assumption could not be properly investigated. Hereafter, the assumption has been made that the Markov assumption applied. Investigation of the proportional hazards assumptions showed a violation only for the variable ICV (transition 1), with a p value of 0.01.

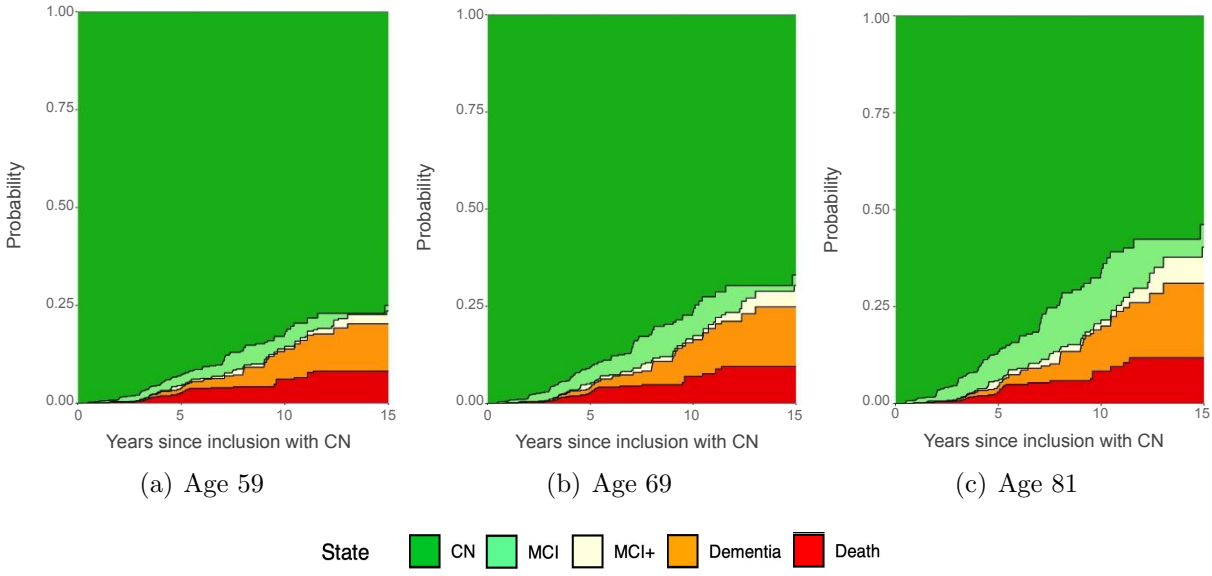
Table 4: Model outcomes for the multi-state model starting from the cognitively normal state at baseline, on the study scale.

	Transition 1 (n = 119)	**Transition 3** (n = 21)	**Transition 4** (n = 35)	**Transition 5** (n = 18)	**Transition 7** (n = 11)
Gender, male	−0.29 (0.27)	0.83 (0.50)	0.19 (0.41)	−0.88 (0.56)	−0.55 (0.84)
Age	0.04 (0.02)		−0.01 (0.04)	−0.04 (0.06)	
MMSE	0.15 (0.25)	−0.28 (0.53)	0.29 (0.53)		
RAVLT					
Immediate	−1.84 (0.28)*				
Forgetting	0.11 (0.21)				
TRABSCOR	0.49 (0.25)				
FAQ	0.74 (0.61)				
APOE4					
Allele 1	0.39 (0.23)				
Allele 2	0.36 (0.73)				
Ventricles	0.42 (0.24)				
Hippocampus	−0.12 (0.25)				
Fusiform	−0.22 (0.25)				
ICV	−0.09 (0.27)				

*p-value is considered significant.
Statistics are presented as mean (SE), unless indicated differently.
Abbreviations: MMSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time; FAQ = Functional Assessment Questionnaire; APOE = Apolipoprotein E; ICV = Intra-Cranial Volume.

Estimated regression coefficients for the transitions with at least one covariate can be found in Table 4; those marked with an * were found to be statistically significant. Inspection of Table 4 shows significance of merely one of all covariates: RAVLT immediate for the transition between the CN and MCI state, with a p value < 0.001. The HR was found to be 0.23 with 95%-confidence interval $(0.13, 0.40)$. This means that an individual in the upper quantile of RAVLT immediate has a 77% lower hazard to experience the transition from CN to MCI, compared to an individual in the lower quantile of RAVLT immediate, that was included in the study at the same time and had identical values for all other covariates.

After determination of the transition-specific effects, the transition probabilities were estimated. State occupation probabilities for patients aged 59, 69 and 81 are shown in Figures 11(a), 11(b) and 11(c), respectively. Inspection of these plots showed a decreased probability to remain in the CN state as the age increases, with a proportional increase to enter the four other states. The figures showing the state occupation probabilities for gender, MMSE quantile and number of APOE4 alleles can be found in Appendix E. Comparison between the state occupation probabilities of the number of APOE4 alleles, showed a noticeably decreased probability to exit the CN state when having no APOE4 alleles compared to having 1 or 2, suggesting that having no APOE4 alleles is protective against Alzheimer's Disease. Comparison between the two MMSE quantiles barely showed differences between the state occupation probabilities. Similarly, the state occupation probabilities for gender were found to be comparable, with the exception that men showed a higher probability to enter the dementia state, while women were more likely to enter the death state.



(a) Age 59 (b) Age 69 (c) Age 81

Figure 11: Study scale state occupation probabilities for the multi-state model starting from CN for three individuals of different ages, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).

Note that none of the covariates for which a comparison was made between the state occupation probabilities were found to be statistically significantly impact-

ing the experiencing of an event, therefore it was not unexpected that the state occupation probabilities were found to be similar between groups.

Overall, we found only RAVLT immediate to significantly impact the occurring of a transition, with a HR of 0.23 $(0.13, 0.40)$ for the transition between the CN and MCI state. Moreover, state occupation probabilities were found to be similar between groups, with minor differences, although cognitive decline was more likely to occur with an increased age and less likely to occur for individuals with no APOE4 alleles.

4.3.2 Model starting from the CN state (age scale)

The model starting from CN at baseline using the age scale, has an identical transition-structure as the model above, with the same number of events (see Figure 10). Model outcomes can be found in Table 5.

Table 5: Model outcomes for the multi-state model starting from the cognitively normal state at baseline, on the age scale.

	Transition 1 (n = 119)	**Transition 3** (n = 21)	**Transition 4** (n = 35)	**Transition 5** (n = 18)	**Transition 7** (n = 11)
Gender, male	−0.26 (0.26)	0.69 (0.51)	0.24 (0.41)	−0.97 (0.57)	−1.48 (1.17)
MMSE	0.21 (0.25)	−0.21 (0.53)	0.26 (0.53)		
RAVLT					
Immediate	−1.28 (0.28)*				
Forgetting	0.10 (0.21)				
TRABSCOR	0.30 (0.24)				
FAQ	0.50 (0.62)				
APOE4					
Allele 1	0.42 (0.22)				
Allele 2	0.25 (0.73)				
Ventricles	0.23 (0.23)				
Hippocampus	−0.06 (0.25)				
Fusiform	−0.13 (0.25)				
ICV	−0.03 (0.27)				

*p-value is considered significant.
Statistics are presented as mean (SE), unless indicated differently.
Abbreviations: MMSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time; FAQ = Functional Assessment Questionnaire; APOE = Apolipoprotein E; ICV = Intra-Cranial Volume.

Similarly to the model on the study scale, the only regression coefficient that was found to have a statistically significant effect on making the transition between the CN and MCI state was RAVLT immediate, with a HR (95%-confidence interval (CI)) of 0.28 $(0.16, 0.48)$ and a p value < 0.001. As such, an individual in the upper quantile of RAVLT immediate has a 72% lower hazard to experience the transition from CN to MCI, compared to an individual in the lower quantile of RAVLT immediate, of the same age, that furthermore had all identical values for all other covariates. Notice also that the estimated regression coefficients and their standard errors differ minimally between the models on the age and the study scale.

Furthermore, general adherence to the cox assumption was found, with the only exception of the variable RAVLT forgetting. Due to non-convergence during checking of the Markov assumption on the study scale, the Markov assumption was not

investigated for this model and was assumed to hold.

After model fitting, the transition probabilities were estimated. However, use of the `probtrans()`-function resulted in the following warning for all individuals: `Warning! Negative diagonal elements of (I+dA); the estimate may not be meaningful.` As such, interpretation of the state occupation probabilities is limited for this model. The warning is hypothesized to be due to the way the transition probabilities are estimated. Namely, as described in Section 3.3 the diagonal elements of $\hat{H}(x)$ are estimated as

$$\hat{H}_{ii}(x) = 1 - \sum_{\mathcal{S} \ni j \neq i} \hat{H}_{ij}(x).$$

As such, having extreme values in the Z-matrix may cause boundary − i.e. large − values for the $\hat{H}_{ij}(x)$, leading to negative diagonal elements. Since the underlying data, nor the estimation method was controllable, this warning could not be solved.

The state occupation probabilities can be bound in Appendix F. Note that the time on the x-axis now represents biological age, rather than the number of years since inclusion in the study. As expected, the state occupation probabilities for three individuals with different ages at baseline were found to be identical. Furthermore, little difference could be observed between the different MMSE quantiles, as well as between the number of APOE4 alleles. When contrasting the two genders, we found women to have a higher probability than men to end in the death state as they grow older, whereas men have a higher probability to end in the dementia state. These outcomes are comparable to the transition probabilities on the study scale.

Overall, little difference was found between the models on the study and age scale, respectively, with significance only having been found for RAVLT immediate, where having higher scores on this test protects against making the transition between the CN and MCI state.

4.3.3 Model starting from the MCI state (study scale)

Moving on to the model with initial state MCI; the number of events per transition can be found in Figure 12. Here, blue and orange indicate the number of individuals starting and ending in that state, respectively, and black indicates the number of individuals that have undergone the transition.

As compared to the models starting in the CN state, more events were observed per transition. As a result, more transition-specific covariates could be fitted: 26, 24, 2, 1, 12 and 6 covariates for transitions 1 to 6, respectively. Since less covariates were available than allowed for, according to the number of transitions, all covariates were used for transitions 1 and 2. As in the previous models, transition-specific covariates for transitions with limited events were chosen based on theoretical knowledge and convergence. Regression coefficients can be found in Table 6; the HR's with 95%-CI's of the statistically significant covariates can be found in Table 8. In this model, many more variables were found to be statistically significant, as indicated by the asterisk in Table 6. This may be due to a more homogeneous patient population compared to the previous two models.

As before, addition of the entering times to the model provoked convergence problems, leaving adherence to the Markov assumption unverified. Moving forward,

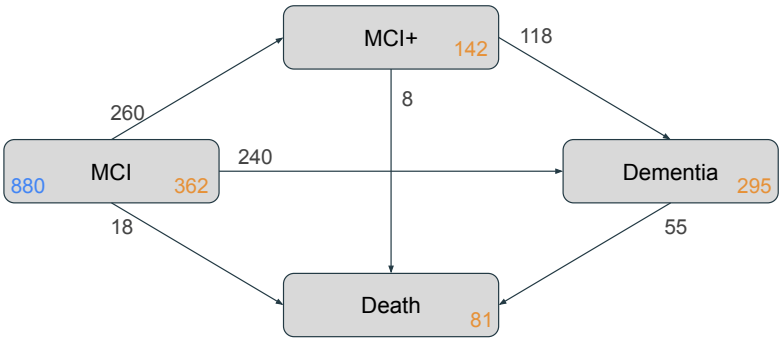


Figure 12: Multi-state model showing the progression to dementia starting from MCI, with blue and orange indicating the number of people starting and ending in a state, and black indicating the number of people undering the transition (MCI = Mild Cognitive Impairment).

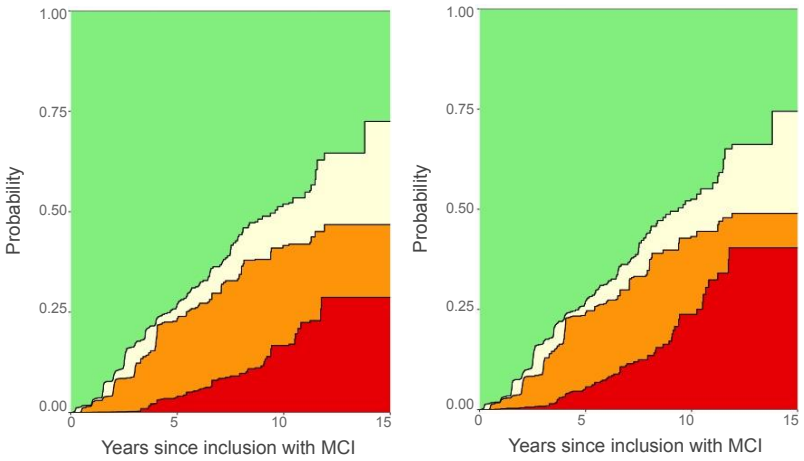
it has been assumed that the Markov assumption holds. Verification of the proportional hazards assumption showed violation for the variable gender in transition 1 and for the variables APOE4, RAVLT forgetting and TRABSCOR in transition 2. As such, the proportional hazards assumption was found to be slightly violated, meaning regression coefficients may have been underestimated. For this application, violation of the proportional hazards assumption was deemed acceptable.

Table 6: Model outcomes for the multi-state model starting from the mild cognitive impairment state at baseline, on the study scale.

	Transition 1 (n = 260)	**Transition 2** (n = 240)	**Transition 3** (n = 18)	**Transition 4** (n = 110)	**Transition 5** (n = 8)	**Transition 6** (n = 55)
Gender, male	0.08 (0.18)	−0.32 (0.19)		0.17 (0.34)	0.00 (0.89)	0.55 (0.38)
Age	0.01 (0.01)	−0.03 (0.01)*	0.11 (0.05)*	0.03 (0.02)		0.05 (0.02)*
MMSE	0.08 (0.15)	−0.21 (0.18)	−0.22 (0.61)	−0.38 (0.28)		−0.77 (0.47)
RAVLT						
Immediate	−0.20 (0.16)	−0.86 (0.18)*		−0.75 (0.31)*		0.41 (0.36)
Forgetting	0.20 (0.15)	0.01 (0.15)		−0.14 (0.25)		−0.44 (0.34)
TRABSCOR	0.28 (0.15)	0.41 (0.16)*		−0.24 (0.28)		
FAQ	0.44 (0.15)*	1.03 (0.16)*		0.36(0.27)		
APOE4						
Allele 1	0.16 (0.16)	0.70 (0.17)*		0.34 (0.29)		−0.01 (0.37)
Allele 2	0.69 (0.23)*	0.90 (0.24)*		0.47 (0.39)		−0.02 (0.44)
Ventricles	0.07 (0.16)	0.36 (0.17)*		−0.81 (0.28)*		
Hippocampus	−0.49 (0.17)*	−0.88 (0.18)*		−0.21 (0.28)		
Fusiform	0.00 (0.16)	−0.33 (0.17)		0.07 (0.31)		
ICV	−0.18 (0.18)	0.23 (0.19)				

*p-value is considered significant.
Statistics are presented as mean (SE), unless indicated differently.
Abbreviations: MMSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time FAQ = Functional Assessment Questionnaire; APOE = Apolipoprotein E; ICV = Intra-Cranial Volume.

Next, the transition probabilities were estimated. For illustration purposes, the state occupation probabilities for two individuals of different ages are shown in Figures 13(a) and 13(b); all other visualisations can be found in Appendix G. In contrast with the many significant effects found in our Cox model, the state occupation probabilities show little difference between groups.



(a) Age 60 (b) Age 72

Figure 13: Study scale state occupation probabilities for the multi-state model starting from MCI for two individuals of different ages, with identical values for all other covariates (MCI = Mild Cognitive Impairment).

In Figures 13(a) and 13(b), it can be seen that a patient aged 72 at baseline has a slightly higher chance of dying rather than developing dementia, compared to a patient aged 60 at baseline, who is more likely to develop dementia rather than dying. Analogously, individuals in the lower quantile for the MMSE variable are more likely to die rather than develop dementia (Figure 22(a)), whereas those in the upper quantile have a higher probability to develop dementia (Figure 22(b)). Figures 23(a) and 23(b) show that individuals with one APOE4 allele have a higher chance of leaving the MCI state to any other state, compared to having no APOE4 alleles. Unfortunately, due to restrictions in the data, no three patients with identical characteristics could be found such that all three counts of APOE4 alleles could be compared. Examination of the state occupation probabilities between men and women (Figures 21(b) and 21(a), respectively) show no noticeable differences.

4.3.4 Model starting from the MCI state (age scale)

The last model, starting from MCI at baseline with time measured on the age scale, uses the same data as the previous model, of which the number of events per transition can be found in Figure 12. Using this data, the model on the age scale was fit using the same selection of transition-specific covariates as for the model described in Section 4.3.3, with exception that the age variable was removed due to it being absorbed into the observed time. The model outcomes can be found in Table 7.

Table 7: Model outcomes for the multi-state model starting from the mild cognitive impairment state at baseline, on the age scale.

	Transition 1 (n = 260)	**Transition 2** (n = 240)	**Transition 3** (n = 18)	**Transition 4** (n = 110)	**Transition 5** (n = 8)	**Transition 6** (n = 55)
Gender, male	0.11 (0.18)	-0.41 (0.19)$^{\Delta}$		**-0.65** (**0.41**)	-0.59 (1.37)	0.61 (0.37)
MMSE	0.04 (0.15)	-0.26 (0.18)	-0.36 (0.69)	-0.39 (0.34)		-0.59 (0.46)
RAVLT						
Immediate	-0.05 (0.16)	-0.79 (0.18)*		-0.96 (0.37)*		0.66 (0.36)
Forgetting	0.21 (0.15)	0.06 (0.15)		-0.42 (0.27)		-0.33 (0.34)
TRABSCOR	0.20 (0.15)	0.33 (0.16)*		-0.04 (0.33)		
FAQ	0.32 (0.15)*	1.01 (0.16)*		0.67(0.34)		
APOE4						
Allele 1	0.14 (0.16)	0.70 (0.18)*		0.62 (0.37)		-0.33 (0.37)
Allele 2	0.62 (0.23)*	0.90 (0.25)*		1.25 (0.46)$^{\Delta}$		-0.13 (0.45)
Ventricles	**-0.04** (**0.16**)	0.33 (0.17)*		**0.22** (**0.33**)$^{\circ}$		
Hippocampus	-0.32 (0.16)*	-0.80 (0.18)*		-0.28 (0.31)		
Fusiform	0.05 (0.16)	-0.31 (0.17)		**-0.44** (**0.33**)		
ICV	-0.18 (0.18)	0.23 (0.19)				

*p-value is considered significant on study and age scale; $^{\Delta}$p-value considered significant only on age scale; $^{\circ}$p-value considered significant only on study scale.
Bold-faced values are those estimates that have changed sign between the study and age scale.
Statistics are presented as mean (SE), unless indicated differently.
Abbreviations: MMSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time
FAQ = Functional Assessment Questionnaire; APOE = Apolipoprotein E; ICV = Intra-Cranial Volume.

What is most interesting about the model outcomes, is that some estimates have changed sign compared to the model on the study scale. This is counter-intuitive, since shifting the time-scale was expected to only impact the size of the effect, not the direction of the effect. This phenomenon occurred for the variable ventricles in transitions 1 and 4 and the variable fusiform in transition 4. Similarly, differences were found in the significance of variables between the two time-scales, however this can be explained due to the standard error changing with the estimate. Hazard ratio's of the significant regression coefficients can be found in Table 8. From this table, we see that none of the statistically significant regression coefficients changed signs between the two models: HR's stay either below one, or above one, within a transition-specific covariate.

Verification of the proportional hazards assumption showed a violation for the variables gender, ventricles and ICV in transition 1 and TRABSCOR in transition 4. Nonetheless, adherence to the assumption was confirmed. As seen in all previous three models, adherence to the Markov assumption could not be checked due to complications with the convergence, leaving the Markov assumption to be an unverified, yet implemented assumption throughout.

Following the assumption checks, the transition probabilities were estimated; their accompanying state occupation probabilities can be found in Appendix H. As seen before during the estimation of the CN model on the age scale, estimation of the transition probabilities for the current model resulted in the warning message `Warning! Negative diagonal elements of (I+dA); the estimate may not be meaningful`, for all individuals. So, the interpretation of these plots may not be reliable; this conclusion is strengthened when comparing the plots on the age to the study scale, where the plots on the age scale show large jumps in the state occupation probabilities. With regards to the differences between groups, we found that

Table 8: Hazard ratio's of the variables found to be statistically significant for the models starting from mild cognitive impairment at baseline.

	Variable	**Study scale**		**Age scale**	
		HR	**95%-CI**	**HR**	**95%-CI**
Transition 1	FAQ	1.55	(1.16, 2.09)	1.38	(1.03, 1.85)
	APOE2	2.00	(1.28, 3.13)	1.86	(1.19, 3.99)
	Hippocampus	0.61	(0.44, 0.85)	0.73	(0.53, 1.00)
Transition 2	Gender	-	-	0.66	(0.46, 0.96)
	RAVLT immediate	0.42	(0.30, 0.60)	0.45	(0.32, 0.64)
	TRABSCOR	1.51	(1.10, 2.08)	1.39	(1.01, 1.91)
	FAQ	2.81	(2.05, 3.85)	2.76	(2.02, 3.77)
	APOE allele 1	2.01	(1.43, 2.81)	2.01	(1.43, 2.85)
	APOE allele 2	2.47	(1.54, 3.96)	2.45	(1.51, 3.99)
	Ventricles	1.44	(1.04, 1.99)	1.40	(1.01, 1.93)
	Hippocampus	0.42	(0.29, 0.60)	0.45	(0.32, 0.64)
Transition 4	RAVLT immediate	0.47	(0.26, 0.87)	0.38	(0.19, 0.78)
	APOE allele 2	-	-	3.48	(1.40, 8.61)
	Ventricles	0.44	(0.26, 0.76)	-	-

Abbreviations: FAQ = Functional Assessment Questionnaire; APOE = Apolipoprotein E; RAVLT = Rey Auditory Verbal Learning Test; TRABSCOR = Trail Making Test Part B Time; HR = Hazard Ratio; CI = Confidence Interval.

individuals with zero APOE4 alleles seemed to be less likely to develop dementia than those with one APOE4 allele (Figures 27(a) and 27(a), respectively). More specifically, the plot for zero APOE4 alleles shows a less steep incline overall, implying that having no APOE4 alleles is protective against leaving the MCI+ state. Comparison within the other variables showed no noticeable differences between groups.

4.4 Discussion of application

With this data illustration, we have shown an application of multi-state models on real-life data. To do so, we have built two separate MSMs, consisting of a five- and four-stage model, respectively. Both models were fitted on the study, as well as the age scale. Fitting of all models showed convergence problems when combining specific variables, hinting at the model structure being too complex for the data. Similarly, addition of the entering times to verify adherence to the Markov assumption lead to non-convergence throughout. Furthermore, considerable differences were found between the models on the study scale and those on the age scale for the model starting from MCI at baseline, which was not in line with what theory would suggest. Especially reversal of the effect on the outcome, i.e. regression parameters being positive on one scale and negative on the other, was an unexpected finding. The changed sign may be due to a number of reasons, under which (i) large differences in ages between participants, leading to significant differences between the two time scales, (ii) small estimated effects with large standard errors, or (iii) errors in the code and/or data. Additionally, the recurring warning on negative diagonal elements remarkably only occurred when estimating the transition probabilities on the age scale, for both starting states. This suggest that something about

the model on the age scale was fundamentally different than the model on the study scale. On the current data, this anomaly could not be explained.

Moreover, the importance of fitting covariates transition-specifically is highlighted in our model outcomes. Namely, for all covariates that were fit on at least 2 transitions within a model, with exception of the variable FAQ in the two MCI models, a sign change of the regression parameter can be observed between transitions. This indicates a reversal of the direction of the effect between transitions. Of course, due to the large standard errors, it cannot be said with certainty that the direction of the effect is equal to the one of the estimated regression parameter. However, this observation does emphasize that variables may act differently for different transitions.

It should furthermore be noted that the models starting in the CN and the MCI states are not only fundamentally different in their design and interpretation, but due to the study set-up, the patient population in the two groups are also incomparable. Namely, individuals that entered the study in the MCI state were proven to be cognitively declining, while there was no proof that individuals who entered the study in the CN state would ever show cognitive decline at all. As such, patients in the MCI model were a-priori more likely to undergo a transition, than those in the CN model. Similarly, since the data of the three phases from the ADNI project were aggregated for this project, it should be kept in mind that patients between phases may not have been comparable.

Unfortunately, due to missingness and convergence problems, the four fitted models show little practical use. However, comparison between the models on the age and study scales within the baseline categories show interesting behaviour in their discordance, which are promising for further research. Additionally, adding time-varying covariates would increase accuracy of the model, possibly leading to a better representation of the disease mechanism. Moreover, since the Markov assumption could not be explicitly confirmed in any model due to convergence problems, one may wonder about the appropriateness of this assumption. As such, it would be our recommendation to further investigate the difference between the age and study scales for these models, use time-varying covariates, and implement a semi- or non-Markovian approach.

5 Discussion

With this thesis, we aimed to give an overview of the relevant theory behind Markovian multi-state models and show an application of this type of model in the field of Alzheimer's research. Four models were fit using data from the Alzheimer's Disease Neuroimaging Initiative*, of which two used CN as starting state in a five-state model, while the other two used MCI as starting state in a four-state model. For each starting state, a model was fit on the age as well as the study scale. All models were assumed to be Markovian and time-inhomogeneous.

Each model showed convergence problems when adding entering times into the Cox regression formula to test the Markov assumption, leaving the assumption unverifiable. Therefore, it is recommended to re-fit each model using the clock reset approach to ensure Markovianity, or use another semi- or non-Markovian approach altogether. Additionally, both models on the age scale resulted in a warning concerning the diagonal elements of the estimated cumulative hazard matrix when estimating the state occupation probabilities. This has been hypothesized to be due to extreme covariate values, although this could not be explicitly confirmed. Since this warning consistently occurred on the age scale, these model outcomes may not be reliable, and it is recommended that a deeper analysis is done on why it occurred. Apart from the warning occurring solely on the age scale, model outcomes between the two time scales were considerably different. This difference was most apparent for the models starting from MCI at baseline, where effects were found to change direction and statistical significance for covariates were not always in agreement between the two time scales.

An explanation for these outcomes may be the limited number of events per transition, which is simultaneously the first limitation of this study. Another limitation of the data is that individuals were enrolled at any stage of illness, leaving the starting states to be incomparable. Therefore, the models on the age scale are most interpretable for this dataset. Unfortunately, due to the recurring warning, these models may not be reliable.

Several adaptations and expansions of the models fitted here are imaginable. Firstly, as touched upon, it would be interesting to add time-variability of the covariates to the model, allowing for incorporation of the longitudinal aspect of the measured variables. This would give a more realistic representation of the progression, and may have useful applications in practice. In theory, a time-varying Cox regression model could be used to estimate the transition intensities, after which these estimated intensities can be used for the estimation of the state occupation probabilities. To our knowledge, no readily available software exists that facilitates such an analysis. Furthermore, it is recommended to combine the models from all starting states into a singular model to most accurately represent all the data at hand.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf [19].

Next to these two additions, it is possible to fit the models differently. For example, while we used a Cox proportional hazards regression model which uses only fixed effects, it may be beneficial to the model outcomes to add random effects as was previously done by [39, 47, 48, 49]. Addition of random effects will account for the dependency structure in the data, which is introduced due to the inclusion of several measurements per patient. Furthermore, instead of fitting a singular Cox model that includes all transition-specific covariate effects, it is possible to fit separate Cox models for each transition. This gives more flexibility with regards to fitting time-varying covariates, and it may be interesting to compare model outcomes using this approach and the model outcomes of the all-encompassing Cox model. Moreover, the transition structure between the states may be changed in a way that is more similar to [25, 39] and [48], where transitions back-and-forth between cognitive ability are allowed and instead of allowing for a transition between dementia and death, these states are considered to be competing and both absorbing.

All in all, we found that time-inhomogeneous Markovian multi-state models using a Cox proportional hazards model are suitable to model the progression of dementia. However, improvements to the models built here are imaginable, which may lead to better interpretability and possibilities for clinical implementation.

References

[1] Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726, 1978.

[2] Per Kragh D. Andersen, Ørnulf Borgan, Richard Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer, 1993.

[3] Giorgos Bakoyannis. Nonparametric tests for transition probabilities in nonhomogeneous markov processes. *Journal of Nonparametric Statistics*, 32(1):131–156, 2020.

[4] Vlad Stefan Barbu, Alex Karagrigoriou, and Andreas Makrides. Semi-markov modelling for multi-state systems. *Methodology and Computing in Applied Probability*, 19:1011–1028, 2017.

[5] Vlad Stefan Barbu and Nicolas Vergne. Reliability and survival analysis for drifting markov models: modeling and estimation. *Methodology and Computing in Applied Probability*, 21:1407–1429, 2019.

[6] Daniel Commenges and Pierre Joly. Multi-state model for dementia, institutionalization, and death. *Communications in Statistics-Theory and Methods*, 33(6):1315–1326, 2004.

[7] David R Cox and David Oakes. *Analysis of Survival Data*. Chapman & Hall, 1984.

[8] Jacobo de Uña-Álvarez and Micha Mandel. Nonparametric estimation of transition probabilities for a general progressive multi-state model under cross-sectional sampling. *Biometrics*, 74(4):1203–1212, 2018.

[9] Liesbeth C. de Wreede, Marta Fiocco, and Hein Putter. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine*, 99:261–274, 2010.

[10] Liesbeth C de Wreede, Marta Fiocco, and Hein Putter. mstate: an r package for the analysis of competing risks and multi-state models. *Journal of statistical software*, 38:1–30, 2011.

[11] Tanujit Dey, Stuart R Lipsitz, Zara Cooper, Quoc-Dien Trinh, Martin Krzywinski, and Naomi Altman. Survival analysis-time-to-event data and censoring. *Nature Methods*, 19(8):906–908, 2022.

[12] Stephen W Duffy, Iris D Nagtegaal, Matthew Wallis, Fay H Cafferty, Nehmat Houssami, Jane Warwick, Prue C Allgood, Olive Kearins, Nancy Tappenden, Emma O'Sullivan, and Gill Lawrence. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *American journal of epidemiology*, 168(1):98–104, 2008.

[13] Sean R Eddy. Hidden markov models. *Current opinion in structural biology*, 6(3):361–365, 1996.

[14] Marta Fiocco, Hein Putter, and Hans C van Houwelingen. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine*, 27(21):4340–4358, 2008.

[15] Geoffrey Grimmett and Dominic Welsh. *Probability: an introduction*. Oxford University Press, 2014.

[16] Torunn Heggland. Estimating transition probabilities for the illness-death model. Master's thesis, University of Oslo, 2015.

[17] Rune Hoff, Hein Putter, Ingrid Sivesind Mehlum, and Jon Michael Gran. Landmark estimation of transition probabilities in non-markov multi-state models with covariates. *Lifetime data analysis*, 25:660–680, 2019.

[18] Philip Hougaard. Multi-state models: a review. *Lifetime data analysis*, 5:239–264, 1999.

[19] Alzheimer's Disease Neuroimaging Initiative. Adni data use agreement. https://ida.loni.usc.edu/collaboration/access/appLicense.jsp;jsessionid=2DF3B2705E0CD02875AC1B46DA338009. Accessed: 2023-06-16.

[20] Pierre Joly, Cécile Durand, Catherine Helmer, and Daniel Commenges. Estimating life expectancy of demented and institutionalized subjects from interval-censored observations of a multi-state model. *Statistical modelling*, 9(4):345–360, 2009.

[21] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[22] John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

[23] John P Klein, Hans C Van Houwelingen, Joseph G Ibrahim, and Thomas H Scheike. *Handbook of survival analysis*. CRC Press Boca Raton, FL:, 2014.

[24] Thomas Kneib and Andrea Hennerfeind. Bayesian semi parametric multi-state models. *Statistical Modelling*, 8(2):169–198, 2008.

[25] RJ Kryscio, FA Schmitt, JC Salazar, MS Mendiondo, and WR Markesbery. Risk factors for transitions from normal to mild cognitive impairment and dementia. *Neurology*, 66(6):828–832, 2006.

[26] Christopher A Lane, John Hardy, and Jonathan M Schott. Alzheimer's disease. *European journal of neurology*, 25(1):59–70, 2018.

[27] Jennifer Le-Rademacher and Xiaofei Wang. Time-to-event data: An overview and analysis considerations. *Journal of Thoracic Oncology*, 16(7):1067–1074, 2021.

[28] Yimei Li and Qiang Zhang. A weibull multi-state model for the dependence of progression-free survival and overall survival. *Statistics in medicine*, 34(17):2497–2513, 2015.

[29] Robson JM Machado, Ardo van den Hout, and Giampiero Marra. Penalised maximum likelihood estimation in multi-state models for interval-censored data. *Computational Statistics & Data Analysis*, 153:107057, 2021.

[30] Luís Meira-Machado, Jacobo de Uña-Álvarez, and Carmen Cadarso-Suárez. Nonparametric estimation of transition probabilities in a non-markov illness-death model. *Lifetime Data Analysis*, 12:325–344, 2006.

[31] Luís Meira-Machado, Jacobo de Uña-Álvarez, and Somnath Datta. Nonparametric estimation of conditional transition probabilities in a non-markov illness-death model. *Computational Statistics*, 30:377–397, 2015.

[32] Farida Mostajabi and Somnath Datta. Nonparametric regression of state occupation, entry, exit, and waiting times with multistate right-censored data. *Statistics in Medicine*, 32(17):3006–3019, 2013.

[33] Alexandra Nießl, Arthur Allignol, Jan Beyersmann, and Carina Mueller. Statistical inference for state occupation and transition probabilities in non-markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*, 2021.

[34] Hein Putter and Marta Fiocco. Lecture slides on multi-state models. Mastermath course on survival analysis, 2023.

[35] Hein Putter, Marta Fiocco, and Ronald B Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, May 2007.

[36] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[37] Lawrence Rabiner and Biinghwang Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[38] Donald A Redelmeier and Sheldon M Singh. Survival in academy award–winning actors and actresses. *Annals of Internal Medicine*, 134(10):955–962, 2001.

[39] Juan C Salazar, Frederick A Schmitt, Lei Yu, Marta M Mendiondo, and Richard J Kryscio. Shared random effects analysis of multi-state markov models: application to a longitudinal study of transitions to dementia. *Statistics in medicine*, 26(3):568–580, 2007.

[40] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2022.

[41] the ADNI team. *ADNIMERGE: Alzheimer's Disease Neuroimaging Initiative*, 2023. R package version 0.0.1.

[42] Terry M Therneau. *A Package for Survival Analysis in R*, 2021. R package version 3.2-13.

[43] Andrew C Titman. Transition probability estimates for non-markov multi-state models. *Biometrics*, 71(4):1034–1041, 2015.

[44] Andrew C Titman and Hein Putter. General tests of the markov property in multi-state models. *Biostatistics*, 23(2):380–396, 2022.

[45] Bella Vakulenko-Lagun, Micha Mandel, and Yair Goldberg. Nonparametric estimation in the illness-death model using prevalent data. *Lifetime data analysis*, 23:25–56, 2017.

[46] Nicolas Vergne. Drifting markov models with polynomial drift and applications to dna sequences. *Statistical applications in genetics and molecular biology*, 7(1), 2008.

[47] Pei Wang, Erin L Abner, Changrui Liu, David W Fardo, Frederick A Schmitt, Gregory A Jicha, Linda J Van Eldik, and Richard J Kryscio. Estimating random effects in a finite markov chain with absorbing states: Application to cognitive data. *Statistica Neerlandica*, 2023.

[48] Shaoceng Wei, Liou Xu, and Richard J Kryscio. Markov transition model to dementia with death as a competing event. *Computational statistics & data analysis*, 80:78–88, 2014.

[49] Lei Yu, Suzanne L Tyas, David A Snowdon, and Richard J Kryscio. Effects of ignoring baseline on modeling transitions from intact cognition to dementia. *Computational statistics & data analysis*, 53(9):3334–3343, 2009.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California [19].

Appendices

A Mental state definitions

This appendix contains the definitions for three recognized mental states, as defined by the ADNI study. The definitions below are cited directly from the ADNI I protocol. This protocol and further information can be found here.

Cognitively normal (controls)

- No memory complaints aside from those common to other normal subjects of that age range;
- Normal memory function documented by scoring at specific cutoffs on the Logical Memory II subscale (delayed paragraph recall) from the Wechsler Memory Scaled - Revised (maximum score is 25):
  - ≥ 9 for 16 or more years of education;
  - ≥ 5 for 8-15 years of education;
  - ≥ 3 for 0-7 years of education;
- Mini-Mental State Exam score between 24 and 30 (inclusive) - exceptions may be made for subjects with < 8 years of education at the discretion of the project director;
- Clinical Dementia Rating is 0, Memory Box score must be 0;
- Cognitively normal, based on absence of significant impairment in cognitive functions or activities of daily living.

Mild cognitive impairment

- Memory complaint by subject or study partner is verified by a study partner;
- Abnormal memory function documented by scoring below the education adjusted cut-off on the Logical Memory II subscale (delayed paragraph recall) from the Wechsler Memory Scaled - Revised (maximum score is 25):
  - ≤ 8 for 16 or more years of education;
  - ≤ 4 for 8-15 years of education;
  - ≤ 2 for 0-7 years of education;
- Mini-Mental State Exam score between 24 and 30 (inclusive) - exceptions may be made for subjects with < 8 years of education at the discretion of the project director;
- Clinical Dementia Rating is 0.5, Memory Box score must ≥ 0.5;

- General cognition and functional performance sufficiently preserved such that a diagnosis of Alzheimer's disease cannot be made by the site physician at the time of the screening visit.

Dementia

- Memory complaint by subject or study partner is verified by a study partner;
- Abnormal memory function documented by scoring below the education adjusted cut-off on the Logical Memory II subscale (delayed paragraph recall) from the Wechsler Memory Scaled - Revised (maximum score is 25):
  - ≤ 8 for 16 or more years of education;
  - ≤ 4 for 8-15 years of education;
  - ≤ 2 for 0-7 years of education;
- Mini-Mental State Exam score between 20 and 26 (inclusive) - exceptions may be made for subjects with < 8 years of education at the discretion of the project director;
- Clinical Dementia Rating is 0.5, Memory Box score must be 1;
- NINCDS/ADRDA criteria for probable AD.

B Inclusion criteria

This appendix contains the inclusion criteria for the ADNI study, as defined by ADNI itself. The criteria below are cited directly from the ADNI I-protocol. This protocol and further information can be found here.

Inclusion criteria:

- Hachinski ≤ 4;
- Age between 55-90;
- Stability of permitted medications for 4 weeks;
- Geriatric Depression Scale < 6;
- Study partner with 10+ hour/week contact, accompanies to visits;
- Visual and auditory acuity adequate for neuropsychological testing;
- Good general health with no diseases preluding enrollment;
- Women must be sterile or two years past childbearing potential;
- Willing and able to complete a 3 year imaging study (2 years for AD subjects);
- At least 6 grades education or work history;
- Must speak English/Spanish fluently;
- Commitment to Neuroimaging and no medical contraindications to MRI;
- Agrees to DNA for ApoE testing and banking;
- Agrees to blood and urine for biomarkers;
- Not enrolled in other trials or studies.

C Variables

This appendix contains explanations of all variables used in the 'adnimerge' dataset from the `ADNIMERGE` package, as obtained from the Alzheimer's Disease Neuroimaging Initiative. Explanations are cited from ADNI, with additions where needed. Variables were categorized based on their usage for the analysis of this thesis.

Outcome variable

- `DX`: Diagnosed mental state (cognitively normal, mild cognitive impairment, dementia).

Identifying variables

- `RID`: Study number;
- `EXAMDATE`: Date of examination.

Study-specific variables

- `COLPROT`: Study protocol of data collection;
- `ORIGPROT`: Original study protocol;
- `PTID`: Original study protocol ID;
- `SITE`: Center at which examination was done;
- `VISCODE`: Visit code;
- `FSVERSION`: FreeSurfer Software Version;
- `FLDSTRENG`: MRI Field Strength;
- `IMAGEUID`: LONI Image ID.

Non-medical patient variables

- `AGE`: Age in years;
- `PTEDUCAT`: Education in years;
- `PTETHCAT`: Ethnicity;
- `PTGENDER`: Gender;
- `PTMARRY`: Marital status;

- `PTRACCAT`: Race.

Medical patient variables

- **Psychological tests**
  - `CDRSB`: Clinical Dementia Rating – Sum of Boxes;
  - `ADAS11`: Alzheimer's Disease Assessment Scale (11 items);
  - `ADAS13`: Alzheimer's Disease Assessment Scale (13 items);
  - `ADASQ4`: Alzheimer's Disease Assessment Scale Delayed Word Recall;
  - `MMSE`: Mini Mental State Examination;
  - `RAVLT.immediate`: Rey Auditory Verbal Learning Test – Immediate;
  - `RAVLT.learning`: Rey Auditory Verbal Learning Test – Learning;
  - `RAVLT.forgetting`: Rey Auditory Verbal Learning Test – Forgetting;
  - `RAVLT.perc.forgetting`: Rey Auditory Verbal Learning Test – Percentage forgetting;
  - `LDELTOTAL`: Logical Memory, Delayed Recall;
  - `DIGITSCOR`: Digit Symbol Substitution Test Score;
  - `TRABSCOR`: Trail Making Test Part B Time;
  - `FAQ`: Functional Assessment Questionnaire;
  - `mPACCdigit`: ADNI modified Preclinical Alzheimer's Cognitive Composite with Digit Symbol Substitution;
  - `mPACCtrailsB`: ADNI modified Preclinical Alzheimer's Cognitive Composite with Trails B;
  - `MOCA`: Montreal Cognitive Assessment.
- **Medical variables**
  - `APOE4`: Number of Apolipoprotein E4 (APOE4) alleles;
  - `FDG`: Fluorodeoxyglucose (FDG)-PET metaROI;
  - `PIB`: Average Pittsburgh compound B (PIB) Standard Uptake Value Ratio (SUVR) of frontal cortex, anterior cingulate, precuneus cortex, and parietal cortex
  - `AV45`: AV45 ratio (cortical grey matter/whole cerebellum) Summary florbetapir cortical SUVR normalized by whole cerebellum;
  - `FBB`: F-florbetaben (FBB) ratio (cortical grey matter/whole cerebellum). Summary florbetaben cortical SUVR normalized by whole cerebellum;
  - `ABETA`: Amyloid-β peptide (biomarker);
  - `TAU`: A microtubule-associated protein (biomarker);

- `PTAU`: Phosphorylated tau (biomarker);
- `EcogPtMem`: Everyday cognition - Participant self report, memory;
- `EcogPtLang`: Everyday cognition - Participant self report, language;
- `EcogPtVisspat`: Everyday cognition - Participant self report, visual/spatial;
- `EcogPtPlan`: Everyday cognition - Participant self report, planning;
- `EcogPtOrgan`: Everyday cognition - Participant self report, organization;
- `EcogPtDivatt`: Everyday cognition - Participant self report, dividing attention;
- `EcogPtTotal`: Everyday cognition - Participant self report, total;
- `EcogSPMem`: Everyday cognition - Study Partner report, memory;
- `EcogSPLang`: Everyday cognition - Study Partner report, language;
- `EcogSPVisspat`: Everyday cognition - Study Partner report, visual/spatial;
- `EcogSPPlan`: Everyday cognition - Study Partner report, planning;
- `EcogSPOrgan`: Everyday cognition - Study Partner report, organization;
- `EcogSPDivatt`: Everyday cognition - Study Partner report, diving attention;
- `EcogSPTotal`: Everyday cognition - Study Partner report, total;
- `Ventricles`: Ventrical volume;
- `Hippocampus`: Hippocampal volume;
- `WholeBrain`: Whole brain volume;
- `Entorhinal`: Entorhinal cortex volume;
- `Fusiform`: Fusiform gyrus volume;
- `MidTemp`: Middle temporal gyrus volume;
- `ICV`: Intra-Crantial Volume.

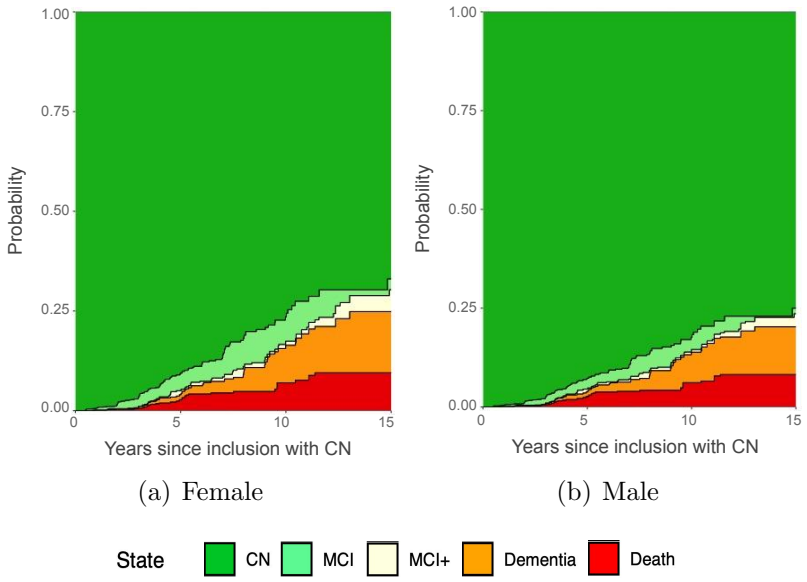
D Percentages missing

Table 9: Missingness for all medical variables, as measured at the moment any new state was entered.

Medical		Psychological	
Variable	**Number missing (%)**	**Variable**	**Number missing (%)**
APOE4	194 (6.7)	CDRSB	380 (13.2)
FDG	1425 (49.4)	ADAS11	424 (14.7)
PIB	2850 (98.8)	ADAS13	438 (15.2)
AV45	1798 (62.3)	ADASQ4	419 (14.5)
FBB	2640 (91.5)	MMSE	416 (14.4)
Amyloid-β	1699 (58.9)	RAVLT	
Tau	1699 (58.9)	Immediate	434 (15.0)
P-tau	1699 (58.9)	Learning	434 (15.0)
EcogPt		Forgetting	438 (15.2)
Memory	1319 (45.7)	Percentage forgetting	445 (15.4)
Language	1318 (45.7)	LDELTOTAL	512 (17.8)
Visual/spatial	1333 (46.2)	DIGITSCOR	1963 (68.1)
Planning	1319 (45.7)	TRABSCOR	504 (17.5)
Organization	1360 (47.2)	FAQ	392 (13.6)
Dividing attention	1333 (46.2)	mPACC	
Total	1316 (45.6)	Digit	419 (14.5)
EcogSP		TrailsB	418 (14.5)
Memory	1312 (45.5)	MOCA	1380 (47.9)
Language	1310 (45.4)		
Visual/spatial	1355 (47.0)		
Planning	1333 (46.2)		
Organization	1383 (48.0)		
Dividing attention	1346 (46.7)		
Total	1312 (45.5)		
Ventricles	660 (22.9)		
Hippocampus	864 (30.0)		
WholeBrain	617 (21.4)		
Entorhinal	911 (31.6)		
Fusiform	911 (31.6)		
MidTemp	911 (31.6)		
ICV	577 (20.0)		

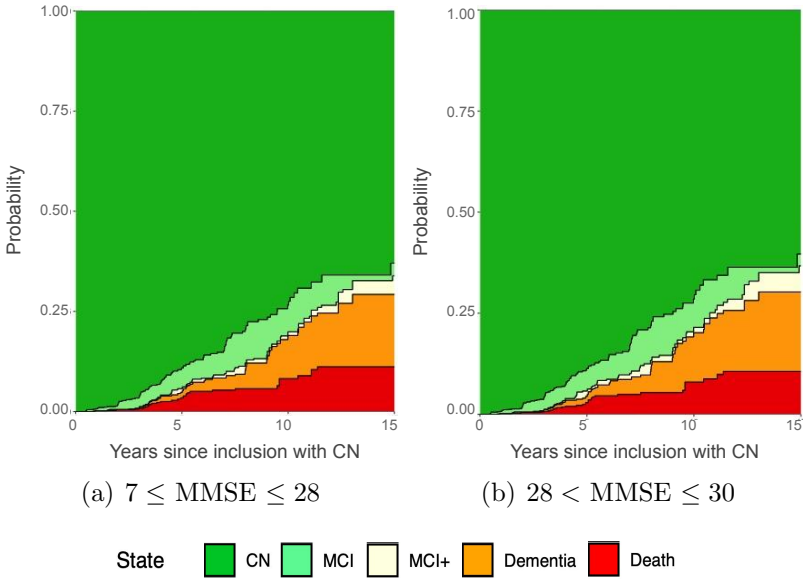
Abbreviations: APOE = Apolipoprotein E; FDG = Fluorodeoxyglucose; PIB = Pittsburgh compound B; FBB = F-florbetaben; EcogPT = Everyday cognition - Participant self report; EcogSTP = Everyday cognition - Study Partner report; ICV = Intra-Cranial Volume; CDRSB = Clinical Dementia Rating – Sum of Boxes; ADAS = Alzheimer's Disease Assessment Scale; MSSE = Mini-Mental State Exam; RAVLT = Rey Auditory Verbal Learning Test; LDELTOTAL = Delayed Total Recall; DIGITSCOR = Digit Symbol Substitution Test Score; TRABSCOR = Trail Making Test Part B Time; FAQ = Functional Assessment Questionnaire; mPACC = Modified Preclinical Alzheimer Cognitive Composite; MOCA = Montreal Cognitive Assessment.

E CN model: state occupation probabilities (study scale)



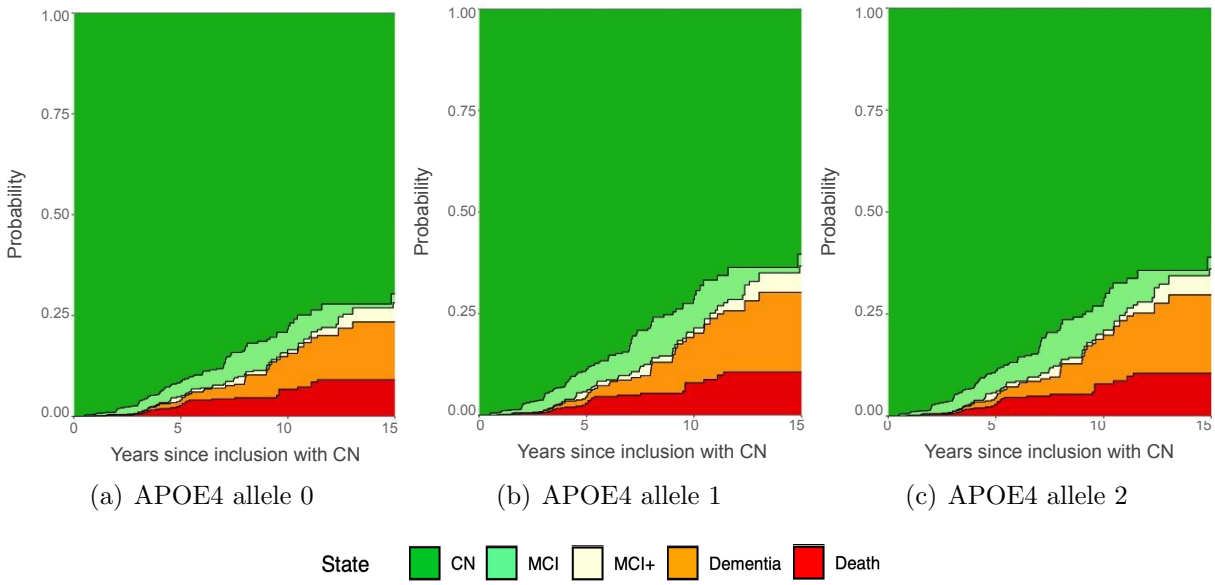
(a) Female (b) Male

Figure 14: Study scale state occupation probabilities for the multi-state model starting from CN for two individuals of different sexes, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).



(a) $7 \leq$ MMSE ≤ 28 (b) $28 <$ MMSE ≤ 30

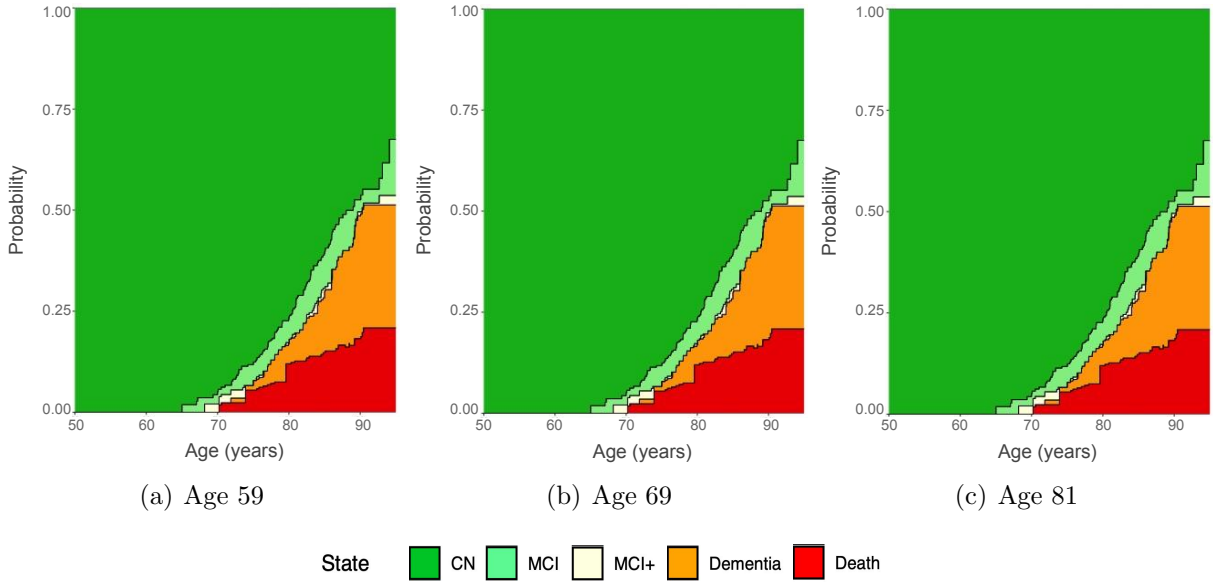
Figure 15: Study scale state occupation probabilities for the multi-state model starting from CN for two individuals in different MMSE quantiles, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).



(a) APOE4 allele 0 (b) APOE4 allele 1 (c) APOE4 allele 2

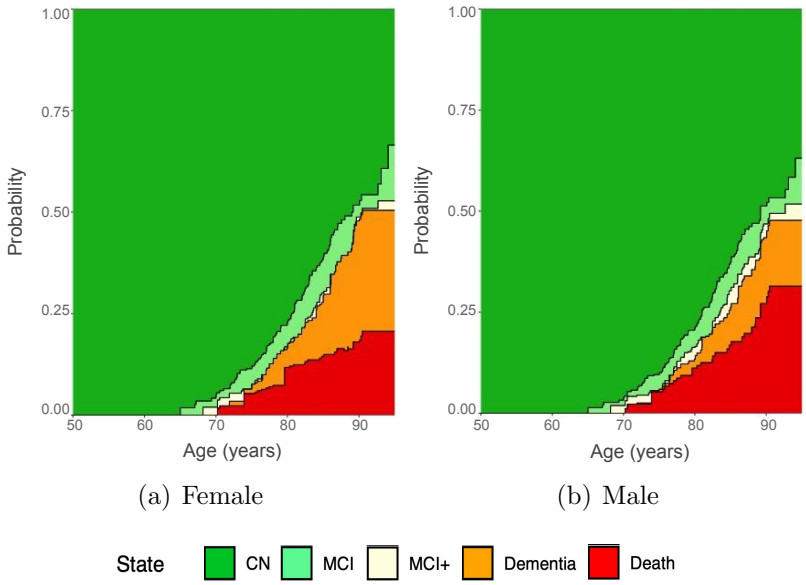
Figure 16: Study scale state occupation probabilities for the multi-state model starting from CN for three individuals with different APOE4 alleles, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment, APOE = Apolipoprotein E).

F CN model: state occupation probabilities (age scale)



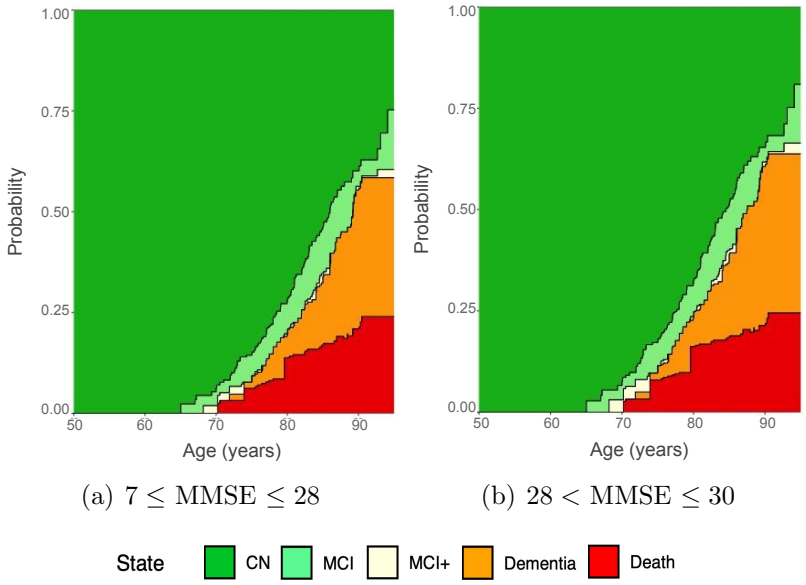
(a) Age 59 (b) Age 69 (c) Age 81

Figure 17: Age scale state occupation probabilities for the multi-state model starting from CN for three individuals of different ages, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).



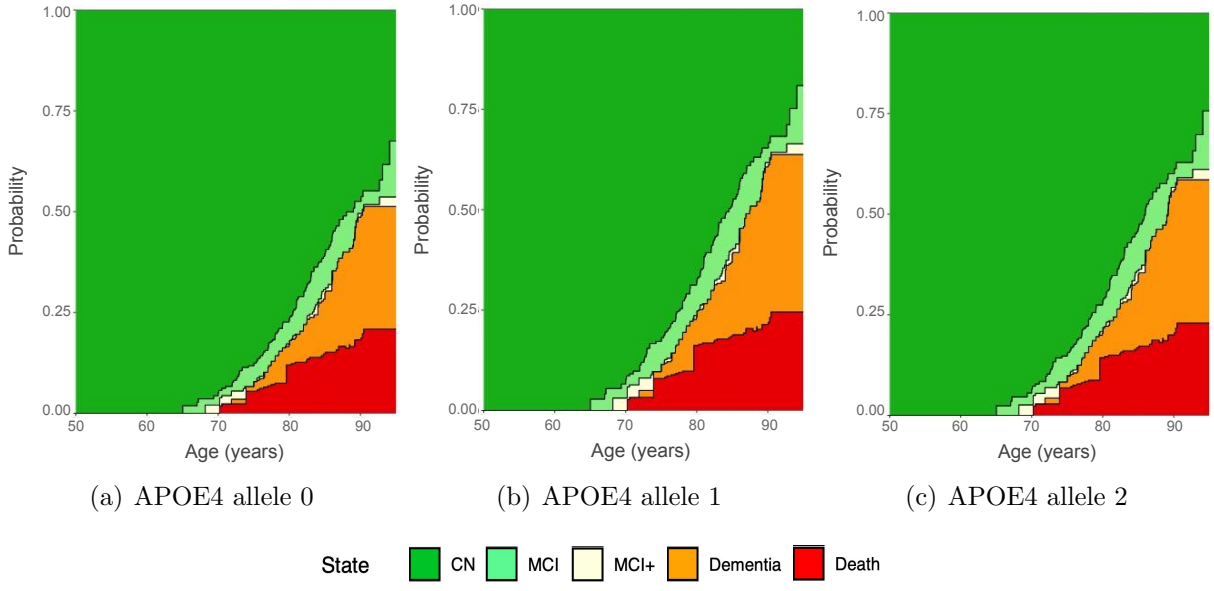
(a) Female (b) Male

Figure 18: Age scale state occupation probabilities for the multi-state model starting from CN for two individuals of different sexes, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).



(a) $7 \leq \text{MMSE} \leq 28$ (b) $28 < \text{MMSE} \leq 30$

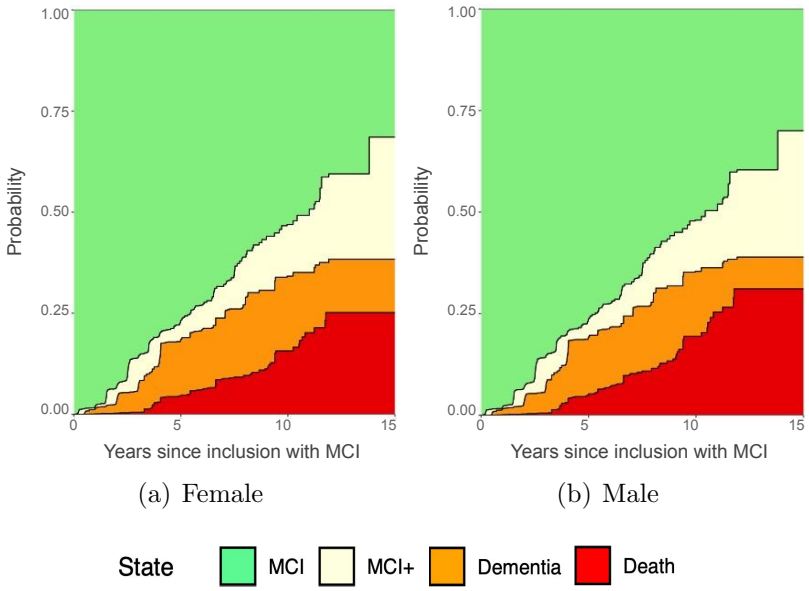
Figure 19: Age scale state occupation probabilities for the multi-state model starting from CN for two individuals in different MMSE quantiles, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment).



(a) APOE4 allele 0 (b) APOE4 allele 1 (c) APOE4 allele 2

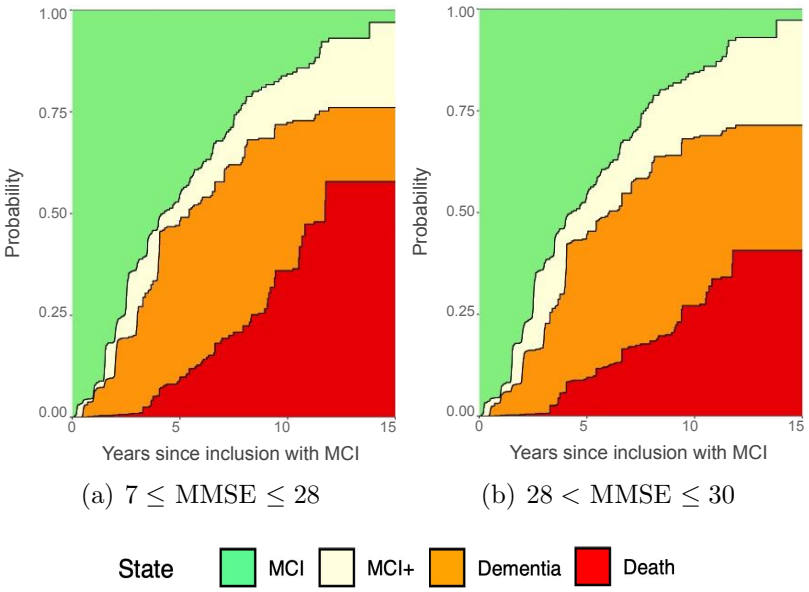
Figure 20: Age scale state occupation probabilities for the multi-state model starting from CN for three individuals with different APOE4 alleles, with identical values for all other covariates (CN = Cognitively Normal, MCI = Mild Cognitive Impairment, APOE = Apolipoprotein E).

G MCI model: state occupation probabilities (study scale)



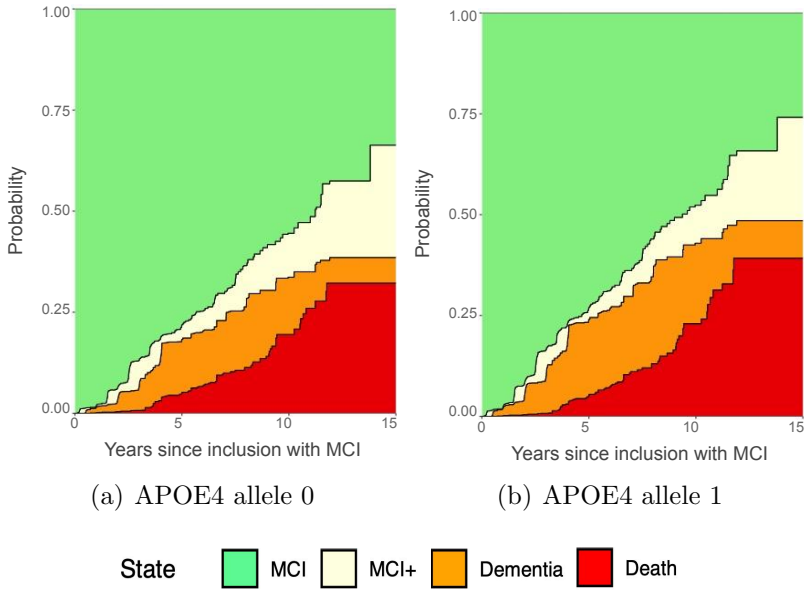
(a) Female (b) Male

Figure 21: Study scale state occupation probabilities for the multi-state model starting from MCI for two individuals of different sexes, with identical values for all other covariates (MCI = Mild Cognitive Impairment).



(a) $7 \leq \text{MMSE} \leq 28$ (b) $28 < \text{MMSE} \leq 30$

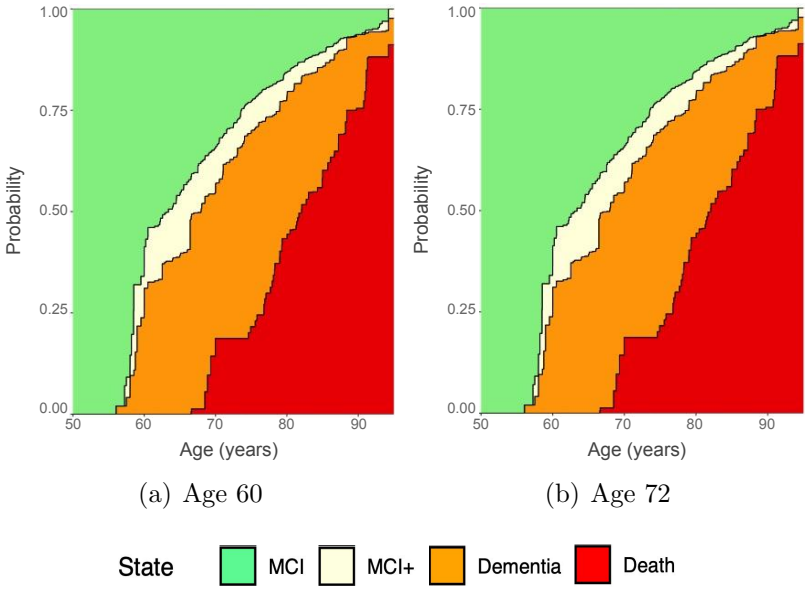
Figure 22: Study scale state occupation probabilities for the multi-state model starting from MCI for two individuals in different MMSE quantiles, with identical values for all other covariates (MCI = Mild Cognitive Impairment).



(a) APOE4 allele 0 (b) APOE4 allele 1

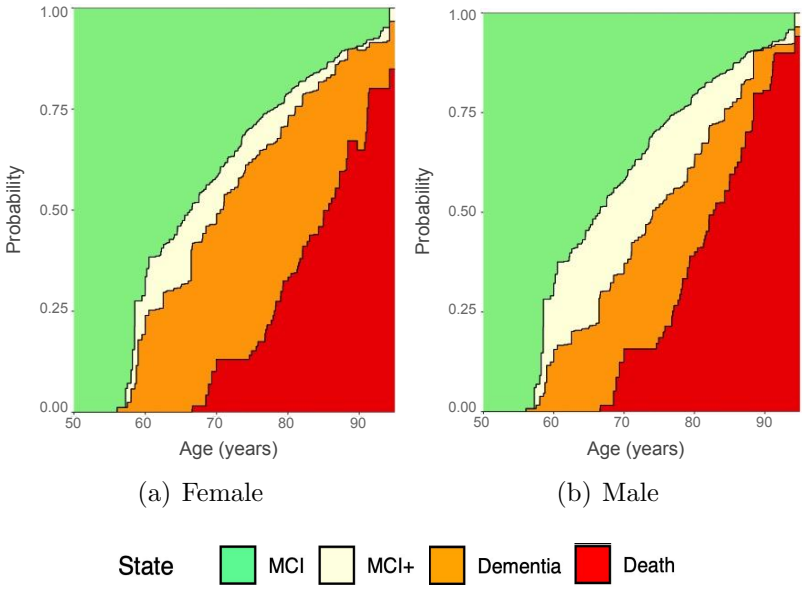
Figure 23: Study scale state occupation probabilities for the multi-state model starting from MCI for two individuals with different APOE4 alleles, with identical values for all other covariates (MCI = Mild Cognitive Impairment).

H MCI model: state occupation probabilities (age scale)



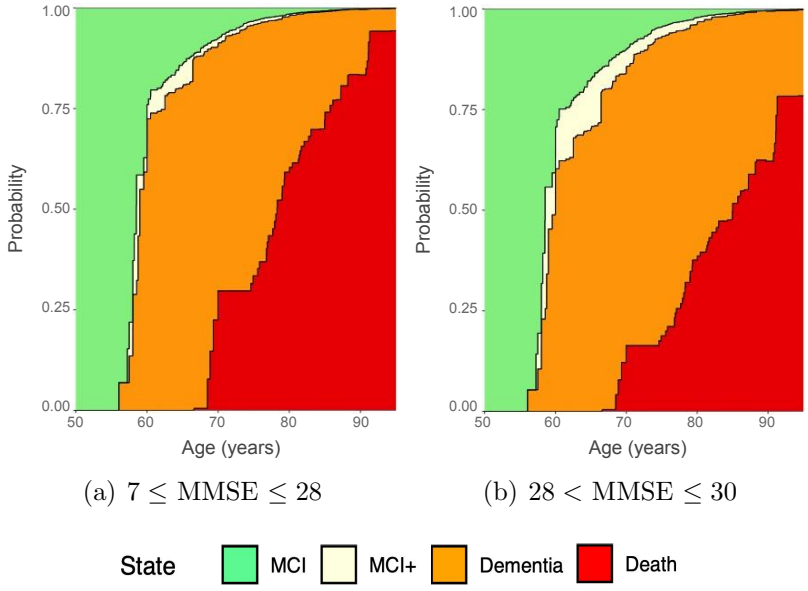
(a) Age 60 (b) Age 72

Figure 24: Age scale state occupation probabilities for the multi-state model starting from MCI for two individuals of different ages, with identical values for all other covariates (MCI = Mild Cognitive Impairment).



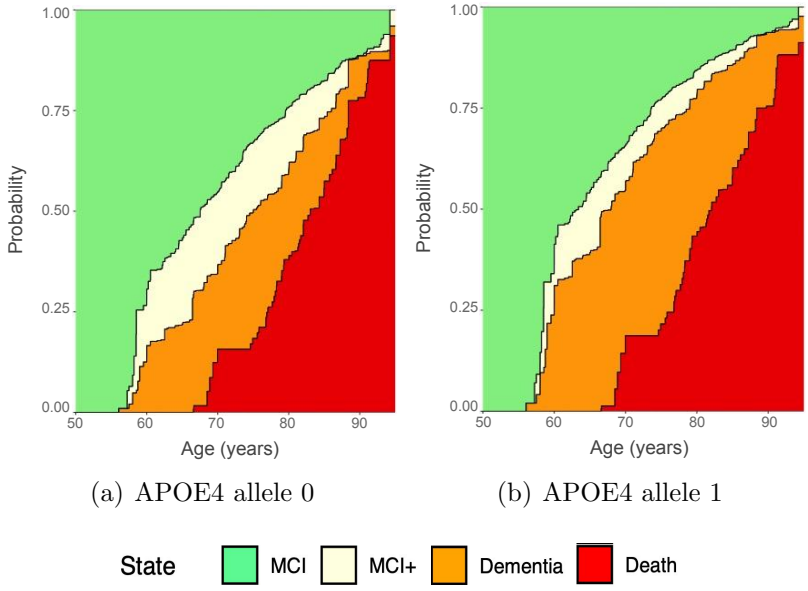
(a) Female (b) Male

Figure 25: Age scale state occupation probabilities for the multi-state model starting from MCI for two individuals of different sexes, with identical values for all other covariates (MCI = Mild Cognitive Impairment).



(a) $7 \leq \text{MMSE} \leq 28$ (b) $28 < \text{MMSE} \leq 30$

Figure 26: Age scale state occupation probabilities for the multi-state model starting from MCI for two individuals in different MMSE quantiles, with identical values for all other covariates (MCI = Mild Cognitive Impairment).



(a) APOE4 allele 0 (b) APOE4 allele 1

Figure 27: Age scale state occupation probabilities for the multi-state model starting from MCI for two individuals with different APOE4 alleles, with identical values for all other covariates (MCI = Mild Cognitive Impairment).