A Multi-state Non-Markov Framework to Estimate Course of Chronic Disease Progression

Ming Ding[1], Haiyi Chen[2], Feng-Chang Lin[2]

[1] Department of Emergency Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2] Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence to:

Ming Ding, ScD

ming_ding@med.unc.edu

Department of Emergency Medicine,

University of North Carolina at Chapel Hill

ABSTRACT

In chronic disease epidemiology, the investigation of disease etiology has largely focused on one endpoint, and the progression of chronic disease as a multi-state process is understudied, representing a knowledge gap. Most existing multi-state regression models require Markov assumption and are unsuitable to describe the course of chronic disease progression that is largely non-memoryless. We propose a new non-Markov framework that allows past states to affect the transition rates of current states, and the key innovation is the conversion of a non-Markov to Markov process by conditioning on past disease history to divide disease states into substates. Specifically, we apply cause-specific Cox models, including past states as covariates, to obtain transition rates of substates, which were used to estimate transition probabilities using the discrete-time Aalen-Johansen estimator. In simulation study, the non-Markov model generated higher coverage rates of transition rates compared to Markov models, particularly for non-Markov process (By applying non-Markov and Markov models, coverage rates were 91% and 88% for Markov process with exponential distribution, 52% and 43% for Markov process with Weibull distribution, 92% and 49% for non-Markov process with exponential distribution, and 59% and 23% for non-Markov process with Weibull distribution). We applied our model to describe the course of coronary heart disease (CHD) progression, where CHD was modeled in healthy, at-risk, CHD, heart failure, and mortality states. In summary, the significance of our framework lies in the fact that transition parameters between disease sub-states may provide a more accurate description of disease course than Markov regression models and shed light on new mechanistic insight into chronic disease. Our method has the potential for wide application in chronic disease epidemiology.

1. INTRODUCTION

In chronic disease epidemiology, investigation of disease etiology has largely focused on one single endpoint. However, the development of chronic disease is a multi-state process, with each state playing a crucial role in affecting its progression. Understanding the course of disease progression allows us to gain new mechanistic insight into the disease at the population level. However, this field is understudied and represents a knowledge gap in epidemiology.

Most existing methods describing the progression of chronic disease are Markov models, which assume that the future state depends only on the present state and not on past history. For the Markov process, transition probabilities can be estimated parametrically from transition rates with constant or Weibull distributions using Kolmogorov differential equations,[1-4] and nonparametrically estimated for other types of distributions using Aalen-Johansen (A-J) estimators.[5-7] To identify factors related to the transition rate, Markov regression models have been proposed that allow the inclusion of covariates.[1, 8] However, they are unsuitable for modeling the progression of chronic diseases that are largely non-memoryless.[9-11] This signifies the urgent need for non-Markov methods to advance the research in estimating the course of disease progression.

Non-Markov models are a largely unexplored field, with only several nonparametric methods proposed.[12-18] Although the A-J estimator can consistently estimate transition probabilities from time 0 to t regardless of Markov assumption,[14, 19, 20] it is systematically biased for any two random time points in the non-Markov process.[16] The landmark Aalen-Johansen (LMAJ) estimator accounts for non-Markov assumption and estimates transition probabilities using landmarking by dividing participants into subgroups according to the state occupation probability at a certain time point.[12] However, the subsampling of participants lowers the power

of estimation, and it is a non-parametrical method which may be difficult to incorporate covariates.[21, 22]

In this paper, we propose a new non-Markov regression framework that allows past states to affect the transition rates of current states, and the key innovation is that we divide disease states into substates to convert non-Markov to Markov process by conditioning on past disease history. Specifically, we apply cause-specific Cox models, including past states as covariates, to obtain transition rates between sub-states, which are used to obtain transition probabilities using a discrete-time A-J estimator.[8]

2. METHODS

A multi-state non-Markov framework. While our method can be used for chronic disease in any number of states, we illustrate our method by assuming five states S_0-S_4 (**Fig 1a**). By conditioning on past states, a disease state can be divided into substates, and a non-Markov process can be converted to a Markov process (**Fig 1b**). Suppose we have survival data with disease states ascertained (**Fig 2a**), which are divided into subsets 1-4 by disease state (**Fig 2b**). Within each subset, we change the data structure from a wide form to a long-form to model transition rates with high flexibility and incorporate time-varying covariates into the framework (**Fig 2c**).

2.1. Estimate transition rate, Q(t). Cause-specific Cox (CSC) models are used to estimate transition rates from one to multiple states. As an extension of the Cox model, CSC assumes different associations of each exposure with each specific event type.[23] To estimate transition rates, we cut time into small intervals in changing data from wide to long-form and model time flexibly, such as using restricted cubic splines. Survival probability by the end of each interval

can be estimated using the 'predict' command of the 'riskRegression' package,[24] and hazard can be estimated as (1-survival probability)/length of the time interval. In particular, for age used as a time scale, the predicted risk at the end of each time interval is approximate to the hazard in discrete time.[24] We fit four CSC models (**Models 1-4**) to data subsets 1-4. The influence of past history of disease on transition rates can be accounted for in models 3 and 4 by including past disease history (as well as their interaction terms with time) as covariates in the model.

Model 1 in subset 1 starting from S_0 state: $\lambda_c(t,X)=\lambda_{0c}(t)\exp(\boldsymbol{\beta}_c + \sum_j \boldsymbol{\beta}_{jc}x_j)$, where c=rate of developing S_1, S_2, S_3, or S_4 states, and x_j are the covariates.

Model 2 in subset 2 starting from S_1 state: $\lambda_c(t,X)=\lambda_{0c}(t)\exp(\boldsymbol{\beta}_c + \sum_j \boldsymbol{\beta}_{jc}x_j)$, where c= rate of developing S_2, S_3, or S_4 states.

Model 3 in subset 3 starting from S_2 state: $\lambda_c(t,X)=\lambda_{0c}(t)\exp(\boldsymbol{\beta}_c + \sum_j \boldsymbol{\beta}_{jc}x_j + \boldsymbol{\beta}_{p1c}x_{p1} + \boldsymbol{\beta}_{p1c:c}x_{p1})$, where c=rate of developing S_3 or S_4 states, and x_{p1} is the past S_1 state.

Model 4 in subset 4 starting from S_3 state: $\lambda_c(t,X)=\lambda_{0c}(t)\exp(\boldsymbol{\beta}_c + \sum_j \boldsymbol{\beta}_{jc}x_j + \boldsymbol{\beta}_{p1c}x_{p1} + \boldsymbol{\beta}_{p2c}x_{p2} + \boldsymbol{\beta}_{p1c:c}x_{p1} + \boldsymbol{\beta}_{p2c:c}x_{p2})$, where c=rate of developing S_4 state, and x_{p1} and x_{p2} are past states S_1 and S_2, respectively.

The estimated hazards are used to construct transition rate matrix **Q(t)** (**Figure 3**), where $\lambda_{01}(t)$, $\lambda_{02}(t)$, $\lambda_{03}(t)$, and $\lambda_{04}(t)$ are the predicted risks at time t from model 1, $\lambda_{12}(t)$, $\lambda_{13}(t)$, and $\lambda_{14}(t)$ are predicted from model 2, $\lambda_{23|0}(t)$, $\lambda_{24|0}(t)$, $\lambda_{23|0_1}(t)$, and $\lambda_{24|0_1}(t)$ are predicted from model 3 with or without history of S_1, $\lambda_{34|0}(t)$, $\lambda_{34|0_1}(t)$, $\lambda_{34|0_2}(t)$, and $\lambda_{34|0_1_2}(t)$ are predicted from model 4 with or without history of S_1 and S_2.

2.2. State occupational probabilities, P(t), are defined as the proportion of participants occupied in each state at t and can be expressed as functions of transition rates. As $\mathbf{P}'(t) = \mathbf{P}(t) * \mathbf{Q}(t)$, $\mathbf{P}(t+1) = \mathbf{P}(t) + \mathbf{P}'(t) = \mathbf{P}(t) + \mathbf{P}(t) * \mathbf{Q}(t) = \mathbf{P}(t) * (I_9 + \mathbf{Q}(t)) = \left(\mathbf{P}(t-1) + \mathbf{P}'(t-1)\right) *$

$\left(I_9 + \mathbf{Q}(\mathrm{t})\right) = \mathbf{P}(\mathrm{t}-1) * (I_9 + \mathbf{Q}(\mathrm{t}-1)) * (I_9 + \mathbf{Q}(\mathrm{t})) = \mathbf{P}(\mathrm{t}=0) * \prod_{\mathrm{t}=0}^{\mathrm{t}=\mathrm{k}}(I_9 + \mathbf{Q}(\mathrm{t}))$, where $\mathbf{P}(\mathrm{t}) = \left(\mathrm{p}_0(\mathrm{t}), \mathrm{p}_1(\mathrm{t}), \mathrm{p}_{2|0}(\mathrm{t}), \mathrm{p}_{2|0_1}(\mathrm{t}), \mathrm{p}_{3|0}(\mathrm{t}), \mathrm{p}_{3|0_1}(\mathrm{t}), \mathrm{p}_{3|0_2}(\mathrm{t}), \mathrm{p}_{3|0_1_2}(\mathrm{t}), \mathrm{p}_4(\mathrm{t})\right)$, $\mathbf{Q}(\mathrm{t})$ is the transition rate matrix, and I_9 is an identity matrix.

2.3. Transition probabilities, TP (k_1, k_2), is defined as the probability of state transition over a period and can be estimated using a discrete-time A-J estimator, where $\mathbf{TP}(\mathrm{k}_1, \mathrm{k}_2) = \prod_{\mathrm{t}=\mathrm{k1}}^{\mathrm{t}=\mathrm{k2}}(\mathrm{I} + \mathbf{Q}(\mathrm{t}))$.

2.4. Use Bootstrap to obtain 95% CIs. We adopted a parametric bootstrapping approach to repeatedly draw transition rates from their corresponding distributions to estimate the variability of the parameter estimate. We repeat the resampling process 1000 times in each estimation by the CSC model. Statistically, the resulting 1000 parameter estimates represent the empirical distribution. They can be summarized with median values and 95% CI defined by 2.5th to 97.5th percentiles of the empirical distribution.

3. SIMULATION STUDY

3.1. Simulation methods. We performed a simulation study to validate the utility of the non-Markov model and compare it to the Markov model. Data were simulated in five states, and four scenarios were considered: Markov or non-Markov processes with exponential or Weibull distribution. Transition rates were simulated as constant for exponential distribution and increasing/decreasing with time for Weibull distribution. Parameters of transition rates differed by past states for the non-Markov process. We simulated survival data starting from each disease state using 'crisk.sim' package in R, and the parameters used for the simulation study are shown in **Table S1**. Processes with exponential distribution were a special case of Weibull distribution,

with the ancillary parameters fixed at 1. We simulated the datasets 1000 times, with a sample size of 5000 for each simulation.

For each scenario, we compared the performances of the non-Markov and Markov models in the transition rates estimation of sub-states. In changing data from wide to long format, we cut time into intervals by 0.05. However, if the Cox model did not converge due to scarce cases within some intervals, we cut it by 0.1. Restricted cubic splines were used to model the event indicator highly flexibly at each time interval. In the non-Markov framework, we included past disease history and the interaction terms with time in models 3-4. Survival probability by the end of each interval was obtained, and the hazard was estimated as (1-survival probability)/length of time interval. We estimated transition rates at time points 0.3, 0.6, 0.9, 1.2, and 1.5 when most simulated participants developed the endpoint. The estimated transition rates were compared to theoretical rates, calculated as

$$\frac{1/\text{ beta0}}{(e^{\text{ancillary parameters}})\text{^}(1/\text{ beta0})} \times t^{(1/\text{ beta0}-\ 1)}.$$

Mean squared error (MSE), coverage, and width of 95% CI were obtained. MSE is the average squared difference between the estimate and true value. Coverage is defined as the proportion of 95% CIs covering the true value over all simulated datasets. We compared the coverage to determine which method was closer to the 95% nominal coverage rate. Width of 95% CI is the difference between 95% upper and lower levels. To compare transition rates across models in a summarized manner, we calculated mean MSE, coverage, and width of 95% CI across all states and all time points.

3.2. Simulation results. In Markov scenarios, both Markov and non-Markov models worked well, as expected (**Table 1**). In non-Markov scenarios, the non-Markov model outperformed the Markov model by showing lower MSE and higher coverage rates. For a non-Markov process

with exponential distribution, the coverage rates were 92% and 49%, respectively, using non-Markov and Markov models. For a non-Markov process with Weibull distribution, the coverage rates were 59% and 23% using the non-Markov and Markov models, respectively. The better performance of non-Markov compared to the Markov model was mainly due to the lower MSE and higher coverage rates in estimated transition rates from states 2 and 3 and transition rates from states 3 to 4, which were affected by past disease history (**Table 2**).

4. APPLICATION STUDY

4.1. Study population. The Atherosclerosis Risk in Communities Study (ARIC) study was designed to investigate the causes of atherosclerosis and its clinical outcomes, as well as variations in cardiovascular risk factors and disease by sex and race.[25] The enrolled participants ARIC underwent a phone interview and clinic visit at baseline and were followed up by telephone calls and re-examinations. Participants were contacted periodically by phone and interviewed about interim hospital admissions, cardiovascular outpatient diagnoses, and deaths. Participants who reported CVD-related events were asked to provide medical records that were reviewed by physicians.[26] For each event, the month and year of diagnosis were recorded as the diagnosis date. Heart failure was ascertained by surveillance calls or clinic visits and was verified from death certificates, medical records, and outpatient diagnoses. Deaths were identified from systematic searches of the vital records of states and of the National Death Index, supplemented by reports from family members and postal authorities.[27] We obtained ARIC data through BioLINCC, an open repository created by NHLBI, with the IRB deemed exempt by the University of North Carolina (UNC)-Chapel Hill Review Board.

4.2. Methods. We modeled coronary heart disease (CHD) progression in five states: Healthy, at risk, CHD, heart failure, and mortality. At-risk state was defined as development of hypertension, hyperlipidemia, or diabetes. Hypertension was defined as blood pressure ≥140/90 mmHg or a history of hypertension or use of blood pressure medications.[28] Hyperlipidemia included primary hypertriglyceridemia (≥175 mg/dL) or primary hypercholesterolemia (LDL-c 160–189 mg/dL, and/or non-HDL-c 190–219 mg/dL).[29] Diabetes was defined as fasting glucose ≥7.7mmol/L.[30] We calculated the time of incidence of at-risk state as the earliest time that any of the risk factors were documented.

We assume a forward model of CHD progression while recognizing that it can be backward (*e.g.*, some participants may develop CHD before at-risk state). Our model can deal with this reverse scenario by treating CHD to at-risk state as a new transition. However, given the low proportion of participants who develop CHD before the at-risk state, including this transition may lower study power. Instead, we classify those following reverse transition to a path of forward transition (*e.g.*, 'healthy → CHD → at risk → mortality' is classified to the path 'healthy → CHD → mortality').

We applied our non-Markov framework to examine the dynamics of CHD progression. Briefly, we divided the longitudinal data into four subsets, "individuals start from healthy state to incidence of at-risk, CHD, heart failure, or mortality, whichever outcome came first," "individuals start from at-risk state to incidence of CHD, heart failure, or mortality," "individuals start from CHD state to incidence of heart failure or mortality," and "individuals start from heart failure state to incidence of mortality." Within each subset, we changed the data into a long format by cutting them into small intervals by age. We modeled age flexibly in cubic terms and included past disease history and the interaction terms with age as covariates in models 3 and 4.

4.3. Results. Our study included 15,027 participants who did not develop CHD or heart failure at baseline. During a median of 27 years of follow-up, our study documented 13,043 at-risk participants, 2565 incident cases of CHD, 3283 incident cases of heart failure, and 7677 incident cases of mortality. The transition rates from healthy, at-risk, or CHD states to subsequent states were low at mid-age and gradually increased with age (**Figure 4**). There was a drastic increase in the transition rate from healthy or at-risk status to mortality after 80 years. The transition rate from heart failure to mortality was higher across all ages compared to transition rates between other states. For the transition from heart failure to mortality, participants with a disease history had a higher rate of mortality than those without, with the highest transition rate observed among participants with a history of risk factors and CHD. We provided transition rates between substates at ages 70 and 90 years in **Table 3**.

We estimated the transition probability from each state to the following states starting from 45 years old (**Figure 5**). Participants who started from a healthy state were very likely to develop at-risk state over follow-up. The transition probability from healthy to at-risk state first increased and then decreased due to participants' transition to next states. The risk of mortality was higher for participants who had CHD at 45 years old with a history of risk factors than those without. For participants who had heart failure at 45 years old, the risk of mortality was greater than 90% at age 70 years without disease history and even higher for those with a history of risk factors and CHD. We provided transition probabilities between substates at ages 70 and 90 years in **Table 4**.

We estimated state occupational probability for participants starting from healthy states at age 45. The proportion of participants in the healthy state decreased over time, and the proportion in the mortality state increased over time (**Figure 6**). The proportion of at-risk

participants first increased and then decreased because these participants moved on to the next states, such as CHD, heart failure, or mortality states. At each age, the proportions of participants in CHD or heart failure states were low, either due to their low incidences from previous states or their high risks of moving to the next states. For example, the proportion of participants who were in healthy, at-risk, CHD, heart failure, and mortality states at age 90 years were 1.17, 13.28, 1.85, 6.12, and 77.47%, respectively (**Table 5**). Most of the participants in CHD or heart failure states had a history of risk factors: 87% of CHD participants had risk factors, and 82% of heart failure participants had risk factors or CHD history.

5. DISCUSSION

We proposed a new non-Markov regression model that allows past states to affect current states, and the innovation lies in that a non-Markov process is converted to a Markov process by conditioning on past states and dividing disease states into substates. The simulation study showed that our method generated less biased transition rates for the non-Markov process compared to Markov regression models, which was observed for both exponential and Weibull distributions. In the application study, we applied our model to describe the course of CHD progression.

As the leading cause of mortality in the U.S.,[31] the etiology and prevention of CHD have been investigated for decades and studies have demonstrated that CHD is largely preventable, with modifiable risk factors accounting for 82% of CHD events,[32] including smoking,[33] body mass index (BMI),[34-36] diet quality,[37-39] and physical activity.[40, 41] However, these studies largely focused on one single endpoint, such as incidences of hypertension, CHD, or heart failure.[42-47] A paucity of studies examined CHD progression in small sample sizes using Markov models,

leading to imprecise estimation and limited generalizability of the findings.[48-51] Markov models are unsuitable to describe CHD progression for two reasons. First, CHD progression is largely non-memoryless: Once a person is diagnosed with CHD-related disease, that diagnosis typically persists. Second, the Markov model is memoryless and thus does not allow for multimorbidity, a common condition that accelerates CHD progression and shortens life expectancy.[9-11]

The implication of our application study lies in two aspects. First, the estimated age-specific transition rates indicate disease mechanisms and can guide disease state-specific precision strategies targeting participants at risk of CHD. For example, compared to those without past disease history, participants with past disease history were more likely to transit from one state to another, showing the importance of disease control in the early stages of the progression. With high transition rates from heart failure to mortality observed across all ages, our study highlighted the urgency of prevention of mortality for heart failure patients starting from mid-age. Second, life expectancy in the U.S. has drastically increased over the past 50 years, primarily driven by a continuous increase in survival rate after CVD prognosis.[52, 53] This has led to increased demand for health services and high costs for treatment and residential care. The state occupational probabilities of disease states over time can provide a more accurate estimation of disease burden and facilitate medical resources planning for real-time. Moreover, the ARIC data has unique advantages for our application, including the long-term follow-up to capture the CHD course, CHD endpoints ascertained by medical records, and access to the data through an NHLBI data repository, BioLINCC.

In summary, we developed a multi-state, non-Markov framework that allows past states to affect the transition rates of current states. The transition parameters between substates may enhance our understanding of new mechanisms of disease progression and stimulate new

approaches to disease prevention. Our framework is highly suitable for estimating the progression of non-memoryless chronic diseases and has potentially broad applications in chronic disease epidemiology.

ACKNOWLEDGMENTS Dr. Ming Ding is supported by National Institutes of Health R21 HG012365, and Dr. Feng-Chang Lin is supported by National Center for Advancing Translational Sciences (NCATS) UM1TR004406.

DATA AVAILABILITY STATEMENT The R code for the non-Markov framework is available at the GitHub repository (https://github.com/mingding-hsph/A-non-Markov-regression-framework).

Reference

1. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C, Andersen PK. Multi-state models for the analysis of time-to-event data. Stat Methods Med Res. 2009;18(2):195-222. Epub 20080618. doi: 10.1177/0962280208092301. PubMed PMID: 18562394; PMCID: PMC2692556.
2. Perez-Ocon R, Ruiz-Castro JE, Gamiz-Perez ML. A piecewise Markov process for analysing survival from breast cancer in different risk groups. Stat Med. 2001;20(1):109-22. Epub 2001/01/03. doi: 10.1002/1097-0258(20010115)20:1<109::aid-sim615>3.0.co;2-n. PubMed PMID: 11135351.
3. Saint-Pierre P, Combescure C, Daures JP, Godard P. The analysis of asthma control under a Markov assumption with use of covariates. Stat Med. 2003;22(24):3755-70. Epub 2003/12/16. doi: 10.1002/sim.1680. PubMed PMID: 14673936.

4. Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. Stat Med. 2002;21(22):3369-82. Epub 2002/10/31. doi: 10.1002/sim.1277. PubMed PMID: 12407678.
5. Aalen OO, Johansen S. An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations. . Scandinavian Journal of Statistics. 1978;5:141-50.
6. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53:457-81.
7. Strauss D, Shavelle R. An extended Kaplan-Meier estimator and its applications. Stat Med. 1998;17(9):971-82. doi: 10.1002/(sici)1097-0258(19980515)17:9<971::aid-sim786>3.0.co;2-q. PubMed PMID: 9612885.
8. de Wreede LC, Fiocco M, Putter H. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. Comput Methods Programs Biomed. 2010;99(3):261-74. Epub 2010/03/17. doi: 10.1016/j.cmpb.2010.01.001. PubMed PMID: 20227129.
9. Drozd M, Relton SD, Walker AMN, Slater TA, Gierula J, Paton MF, Lowry J, Straw S, Koshy A, McGinlay M, Simms AD, Gatenby VK, Sapsford RJ, Witte KK, Kearney MT, Cubbon RM. Association of heart failure and its comorbidities with loss of life expectancy. Heart. 2021;107(17):1417-21. Epub 20201105. doi: 10.1136/heartjnl-2020-317833. PubMed PMID: 33153996; PMCID: PMC8372397.
10. Khan MS, Samman Tahhan A, Vaduganathan M, Greene SJ, Alrohaibani A, Anker SD, Vardeny O, Fonarow GC, Butler J. Trends in prevalence of comorbidities in heart failure clinical trials. Eur J Heart Fail. 2020;22(6):1032-42. Epub 2020/04/16. doi: 10.1002/ejhf.1818. PubMed PMID: 32293090; PMCID: PMC7906002.
11. Tran J, Norton R, Conrad N, Rahimian F, Canoy D, Nazarzadeh M, Rahimi K. Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: A population-based cohort study. PLoS Med. 2018;15(3):e1002513. Epub 20180306. doi: 10.1371/journal.pmed.1002513. PubMed PMID: 29509757; PMCID: PMC5839540.
12. Putter H, Spitoni C. Nonparametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. Stat Methods Med Res. 2018;27(7):2081-92. Epub 2018/05/31. doi: 10.1177/0962280216674497. PubMed PMID: 29846146.
13. de Una-Alvarez J, Meira-Machado L. Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. Biometrics. 2015;71(2):364-75. Epub 2015/03/05. doi: 10.1111/biom.12288. PubMed PMID: 25735883.
14. Pepe MS. Inference for Events With Dependent Risks in Multiple Endpoint Studies. Journal of the American Statistical Association. 1991;86(415):770-8.
15. Andersen PK, Wandall ENS, Pohar Perme M. Inference for transition probabilities in non-Markov multi-state models. Lifetime Data Anal. 2022;28(4):585-604. Epub 20220628. doi: 10.1007/s10985-022-09560-w. PubMed PMID: 35764854.
16. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. Lifetime Data Anal. 2006;12(3):325-44. Epub 2006/08/19. doi: 10.1007/s10985-006-9009-x. PubMed PMID: 16917736.

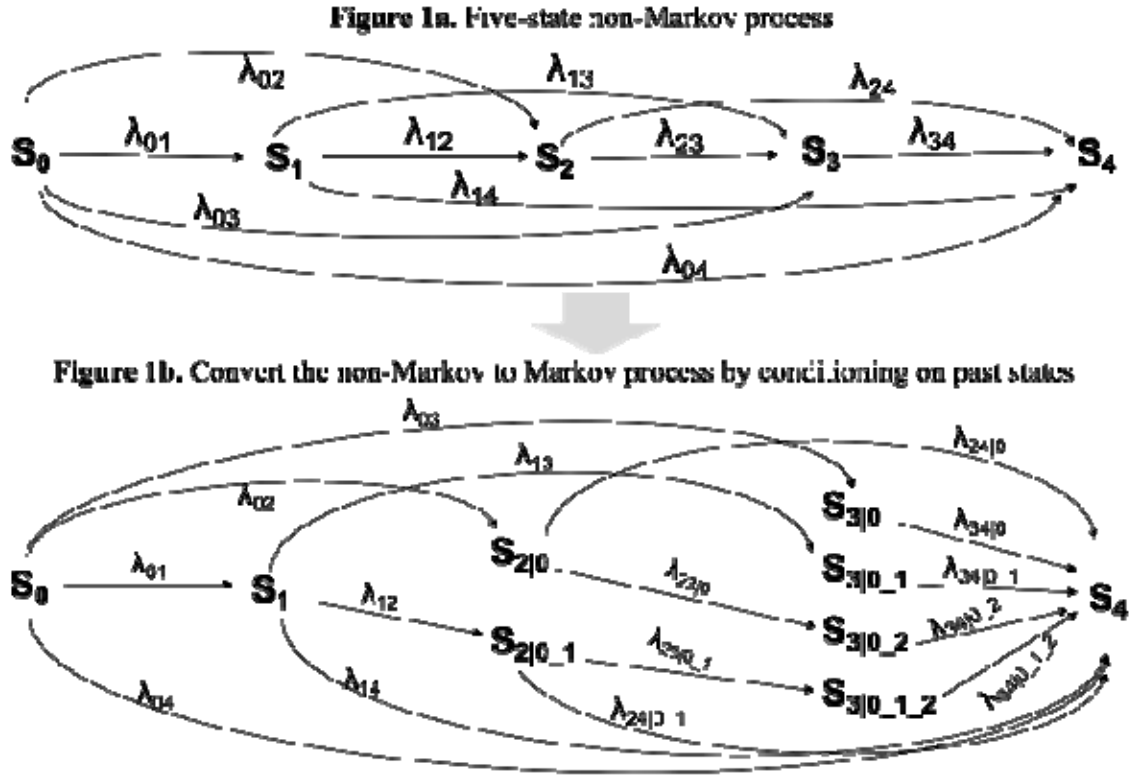
17. Glidden DV. Robust inference for event probabilities with non-Markov event data. Biometrics. 2002;58(2):361-8. doi: 10.1111/j.0006-341x.2002.00361.x. PubMed PMID: 12071409.
18. Meira-Machado L, Sestelo M. Estimation in the progressive illness-death model: A nonexhaustive review. Biom J. 2019;61(2):245-63. Epub 20181120. doi: 10.1002/bimj.201700200. PubMed PMID: 30457674.
19. Datta S, Satten GA. Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. Statistics & Probability Letters. 2001;55(4):403-11.
20. Dattaa S, Sattenb GA. Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. Statistics & Probability Letters. 2001;55 403 – 11.
21. Maltzahn N, Hoff R, Aalen OO, Mehlum IS, Putter H, Gran JM. A hybrid landmark Aalen-Johansen estimator for transition probabilities in partially non-Markov multi-state models. Lifetime Data Anal. 2021;27(4):737-60. Epub 20210930. doi: 10.1007/s10985-021-09534-4. PubMed PMID: 34595580; PMCID: PMC8536588.
22. Hoff R, Putter H, Mehlum IS, Gran JM. Landmark estimation of transition probabilities in non-Markov multi-state models with covariates. Lifetime Data Anal. 2019;25(4):660-80. Epub 20190417. doi: 10.1007/s10985-019-09474-0. PubMed PMID: 30997582.
23. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009;170(2):244-56. Epub 20090603. doi: 10.1093/aje/kwp107. PubMed PMID: 19494242; PMCID: PMC2732996.
24. Gerds TA, Ohlendorff JS, Blanche P, Mortensen R., Wright M., Tollenaar N., Muschelli J., Mogensen U.B., B. O. Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks. https://githubcom/tagteam/riskRegression. 2023.
25. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129(4):687-702. Epub 1989/04/01. PubMed PMID: 2646917.
26. Rose GA. Cardiovascular survey methods. Geneva Albany, N.Y.: World Health Organization; WHO Publications Centre distributor1982.
27. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Archives of internal medicine. 2008;168(7):713-20. Epub 2008/04/17. doi: 10.1001/archinte.168.7.713. PubMed PMID: 18413553.
28. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, Ramirez A, Schlaich M, Stergiou GS, Tomaszewski M, Wainford RD, Williams B, Schutte AE. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. Hypertension. 2020;75(6):1334-57. Epub 20200506. doi: 10.1161/HYPERTENSIONAHA.120.15026. PubMed PMID: 32370572.
29. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, Braun LT, de Ferranti S, Faiella-Tommasino J, Forman DE, Goldberg R, Heidenreich PA, Hlatky MA, Jones DW, Lloyd-Jones D, Lopez-Pajares N, Ndumele CE, Orringer CE, Peralta CA, Saseen JJ, Smith SC, Jr., Sperling L, Virani SS, Yeboah J. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.

Circulation. 2019;139(25):e1046-e81. Epub 20181110. doi: 10.1161/CIR.0000000000000624. PubMed PMID: 30565953.
30. American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33 Suppl 1(Suppl 1):S62-9. doi: 10.2337/dc10-S062. PubMed PMID: 20042775; PMCID: PMC2797383.
31. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, Boehme AK, Buxton AE, Carson AP, Commodore-Mensah Y, Elkind MSV, Evenson KR, Eze-Nliam C, Ferguson JF, Generoso G, Ho JE, Kalani R, Khan SS, Kissela BM, Knutson KL, Levine DA, Lewis TT, Liu J, Loop MS, Ma J, Mussolino ME, Navaneethan SD, Perak AM, Poudel R, Rezk-Hanna M, Roth GA, Schroeder EB, Shah SH, Thacker EL, VanWagner LB, Virani SS, Voecks JH, Wang NY, Yaffe K, Martin SS. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation. 2022;145(8):e153-e639. Epub 20220126. doi: 10.1161/CIR.0000000000001052. PubMed PMID: 35078371.
32. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. N Engl J Med. 2000;343(1):16-22. doi: 10.1056/NEJM200007063430103. PubMed PMID: 10882764.
33. Willett WC, Green A, Stampfer MJ, Speizer FE, Colditz GA, Rosner B, Monson RR, Stason W, Hennekens CH. Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. N Engl J Med. 1987;317(21):1303-9. Epub 1987/11/19. doi: 10.1056/NEJM198711193172102. PubMed PMID: 3683458.
34. Rimm EB, Stampfer MJ, Giovannucci E, Ascherio A, Spiegelman D, Colditz GA, Willett WC. Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men. Am J Epidemiol. 1995;141(12):1117-27. doi: 10.1093/oxfordjournals.aje.a117385. PubMed PMID: 7771450.
35. Bogers RP, Bemelmans WJ, Hoogenveen RT, Boshuizen HC, Woodward M, Knekt P, van Dam RM, Hu FB, Visscher TL, Menotti A, Thorpe RJ, Jr., Jamrozik K, Calling S, Strand BH, Shipley MJ, Investigators B-CC. Association of overweight with increased risk of coronary heart disease partly independent of blood pressure and cholesterol levels: a meta-analysis of 21 cohort studies including more than 300 000 persons. Arch Intern Med. 2007;167(16):1720-8. doi: 10.1001/archinte.167.16.1720. PubMed PMID: 17846390.
36. Mongraw-Chaffin ML, Peters SAE, Huxley RR, Woodward M. The sex-specific association between BMI and coronary heart disease: a systematic review and meta-analysis of 95 cohorts with 1.2 million participants. Lancet Diabetes Endocrinol. 2015;3(6):437-49. Epub 20150507. doi: 10.1016/S2213-8587(15)00086-8. PubMed PMID: 25960160; PMCID: PMC4470268.
37. Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. Am J Clin Nutr. 2000;72(4):912-21. Epub 2000/09/30. doi: 10.1093/ajcn/72.4.912. PubMed PMID: 11010931.
38. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, Gomez-Gracia E, Ruiz-Gutierrez V, Fiol M, Lapetra J, Lamuela-Raventos RM, Serra-Majem L, Pinto X, Basora J, Munoz MA, Sorli JV, Martinez JA, Martinez-Gonzalez MA, Investigators PS. Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med. 2013;368(14):1279-90. Epub 20130225. doi: 10.1056/NEJMoa1200303. PubMed PMID: 23432189.
39. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, Gomez-Gracia E, Ruiz-Gutierrez V, Fiol M, Lapetra J, Lamuela-Raventos RM, Serra-Majem L, Pinto X, Basora J,

Munoz MA, Sorli JV, Martinez JA, Fito M, Gea A, Hernan MA, Martinez-Gonzalez MA, Investigators PS. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. N Engl J Med. 2018;378(25):e34. Epub 20180613. doi: 10.1056/NEJMoa1800389. PubMed PMID: 29897866.
40. Sattelmair J, Pertman J, Ding EL, Kohl HW, 3rd, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease: a meta-analysis. Circulation. 2011;124(7):789-95. Epub 20110801. doi: 10.1161/CIRCULATIONAHA.110.010710. PubMed PMID: 21810663; PMCID: PMC3158733.
41. Tanasescu M, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Exercise type and intensity in relation to coronary heart disease in men. JAMA. 2002;288(16):1994-2000. doi: 10.1001/jama.288.16.1994. PubMed PMID: 12387651.
42. Ding M, Fitzmaurice GM, Arvizu M, Willett WC, Manson JE, Rexrode KM, Hu FB, Chavarro JE. Associations between patterns of modifiable risk factors in mid-life to late life and longevity: 36 year prospective cohort study. BMJ Med. 2022;1(1):e000098. Epub 20220926. doi: 10.1136/bmjmed-2021-000098. PubMed PMID: 36936601; PMCID: PMC9978682.
43. Hu Y, Ding M, Sampson L, Willett WC, Manson JE, Wang M, Rosner B, Hu FB, Sun Q. Intake of whole grain foods and risk of type 2 diabetes: results from three prospective cohort studies. BMJ. 2020;370:m2206. Epub 20200708. doi: 10.1136/bmj.m2206. PubMed PMID: 32641435; PMCID: PMC7341349.
44. Ding M, Li J, Qi L, Ellervik C, Zhang X, Manson JE, Stampfer M, Chavarro JE, Rexrode KM, Kraft P, Chasman D, Willett WC, Hu FB. Associations of dairy intake with risk of mortality in women and men: three prospective cohort studies. BMJ. 2019;367:l6204. Epub 20191127. doi: 10.1136/bmj.l6204. PubMed PMID: 31776125; PMCID: PMC6880246.
45. Ding M, Ellervik C, Huang T, Jensen MK, Curhan GC, Pasquale LR, Kang JH, Wiggs JL, Hunter DJ, Willett WC, Rimm EB, Kraft P, Chasman DI, Qi L, Hu FB, Qi Q. Diet quality and genetic association with body mass index: results from 3 observational studies. Am J Clin Nutr. 2018;108(6):1291-300. doi: 10.1093/ajcn/nqy203. PubMed PMID: 30351367; PMCID: PMC6290366.
46. Ding M, Bhupathiraju SN, Satija A, van Dam RM, Hu FB. Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies. Circulation. 2014;129(6):643-59. Epub 20131107. doi: 10.1161/CIRCULATIONAHA.113.005925. PubMed PMID: 24201300; PMCID: PMC3945962.
47. Farvid MS, Ding M, Pan A, Sun Q, Chiuve SE, Steffen LM, Willett WC, Hu FB. Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. Circulation. 2014;130(18):1568-78. Epub 20140826. doi: 10.1161/CIRCULATIONAHA.114.010236. PubMed PMID: 25161045; PMCID: PMC4334131.
48. Mohammad MA, Stone GW, Koul S, Olivecrona GK, Bergman S, Persson J, Engstrom T, Frobert O, Jernberg T, Omerovic E, James S, Bergstrom G, Erlinge D. On the Natural History of Coronary Artery Disease: A Longitudinal Nationwide Serial Angiography Study. J Am Heart Assoc. 2022;11(21):e026396. Epub 20221027. doi: 10.1161/JAHA.122.026396. PubMed PMID: 36300820; PMCID: PMC9673651.
49. Puddu PE, Menotti A. Natural history of coronary heart disease and heart disease of uncertain etiology: Findings from a 50-year population study. Int J Cardiol. 2015;197:260-4. Epub 20150619. doi: 10.1016/j.ijcard.2015.06.046. PubMed PMID: 26148769.
50. Meyer ML, Lin FC, Jaensch A, Mons U, Hahmann H, Koenig W, Brenner H, Rothenbacher D. Multi-state models of transitions in depression and anxiety symptom severity

and cardiovascular events in patients with coronary heart disease. PLoS One. 2019;14(3):e0213334. Epub 20190307. doi: 10.1371/journal.pone.0213334. PubMed PMID: 30845176; PMCID: PMC6405099.
51. Hardy ST, Holliday KM, Chakladar S, Engeda JC, Allen NB, Heiss G, Lloyd-Jones DM, Schreiner PJ, Shay CM, Lin D, Zeng D, Avery CL. Heterogeneity in Blood Pressure Transitions Over the Life Course: Age-Specific Emergence of Racial/Ethnic and Sex Disparities in the United States. JAMA Cardiol. 2017;2(6):653-61. doi: 10.1001/jamacardio.2017.0652. PubMed PMID: 28423153; PMCID: PMC5634332.
52. Collaborators GBDCoD. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1151-210. Epub 2017/09/19. doi: 10.1016/S0140-6736(17)32152-9. PubMed PMID: 28919116; PMCID: PMC5605883.
53. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1789-858. Epub 2018/11/30. doi: 10.1016/S0140-6736(18)32279-7. PubMed PMID: 30496104; PMCID: PMC6227754.

Figure 1. Convert non-Markov to Markov process by splitting disease states into substates.



S_{20}: State 2 from state 0; S_{20_1}: State 2 from states 0 and 1; S_{30} State 3 from state 0; S_{30_1} : State 3 from states 0 and 1; S_{30_2}: State 3 from states 0 and 2; $S_{30_1_2}$: State 3 from states 0. 1 and 2.

Figure 2. Data preparation for applying the non-Markov framework.

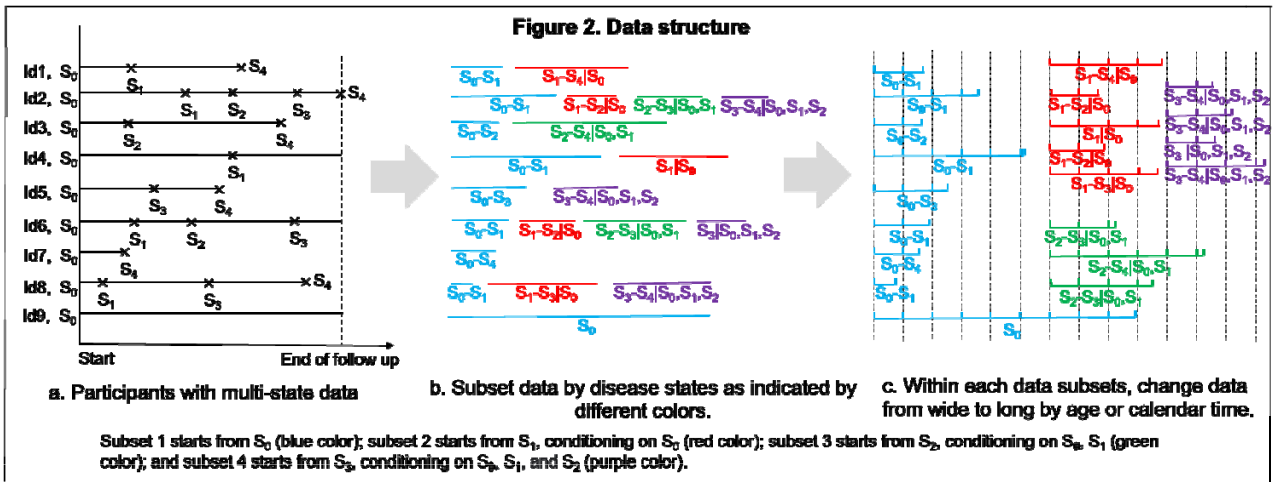


Figure 3. Transition rate matrix Q(t).

Figure 3. Transition rate matrix $\mathbf{Q}(t)$.

$$
\begin{array}{c}
 \\ S_0\,(t) \\ S_1\,(t) \\ S_{2|0}\,(t) \\ S_{2|0_1}\,(t) \\ S_{3|0}\,(t) \\ S_{3|0_1}\,(t) \\ S_{3|0_2}\,(t) \\ S_{3|0_1_2}\,(t) \\ S_4\,(t)
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
S_0\,(t) & S_1\,(t) & S_{2|0}\,(t) & S_{2|0_1}\,(t) & S_{3|0}\,(t) & S_{3|0_1}\,(t) & S_{3|0_2}\,(t) & S_{3|0_1_2}\,(t) & S_4\,(t)
\end{array} \\
\begin{pmatrix}
-(\lambda_{01}+\lambda_{02}+\lambda_{03}+\lambda_{04}) & \lambda_{01} & \lambda_{02} & 0 & \lambda_{03} & 0 & 0 & 0 & \lambda_{04} \\
0 & -\lambda_{12}-\lambda_{13}-\lambda_{14} & 0 & \lambda_{12} & 0 & \lambda_{13} & 0 & 0 & \lambda_{14} \\
0 & 0 & -\lambda_{23|0}-\lambda_{24|0} & 0 & 0 & 0 & \lambda_{23|0} & 0 & \lambda_{24|0} \\
0 & 0 & 0 & -\lambda_{23|0_1}-\lambda_{24|0_1} & 0 & 0 & 0 & \lambda_{23|0_1} & \lambda_{24|0_1} \\
0 & 0 & 0 & 0 & -\lambda_{34|0} & 0 & 0 & 0 & \lambda_{34|0} \\
0 & 0 & 0 & 0 & 0 & -\lambda_{34|0_1} & 0 & 0 & \lambda_{34|0_1} \\
0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{34|0_2} & 0 & \lambda_{34|0_2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{34|0_1_2} & \lambda_{34|0_1_2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

Table 1. A simulation study estimating transition rates (%) compared the non-Markov regression model to Markov model under different scenarios.

	Markov process with exponential distribution		Markov process with Weibull distribution	
	Cox Markov model	Non-Markov model	Cox Markov model	Non-Markov model
MSE	8.88	16.80	60.30	59.46
Width of 95% CI	8.85	13.12	8.93	11.88
Coverage rate	88%	91%	43%	52%
	Non-Markov process with exponential distribution		Non-Markov process with Weibull distribution	
	Cox Markov model	Non-Markov model	Cox Markov model	Non-Markov model
MSE	482.94	17.37	420.89	47.52
Width of 95% CI	6.63	11.25	5.07	10.16
Coverage rate	49%	92%	23%	59%

Mean squared error (MSE) is defined as the average squared difference between estimated value and true value. Width of 95% confidence interval (CI) is defined as the difference between 95% upper and lower levels. Coverage is defined as the proportion of 95% CIs including the true value over all simulated datasets.

Table 2. A simulation study estimating transition rates (%) by disease substates compared the non-Markov regression model to Markov model under different scenarios.

		TR_{01}	TR_{02}	TR_{03}	TR_{04}	TR_{12}	TR_{13}	TR_{14}	$TR_{23\|0}$	$TR_{24\|0}$	$TR_{23\|0_1}$	$TR_{24\|0_1}$	$TR_{34\|0}$	$TR_{34\|0_1}$	$TR_{34\|0_2}$	$TR_{34\|0_1_2}$
Scenarios	Models	MSE														
Markov process with exponential distribution	Markov	1.21	2.15	1.93	1.34	23.75	20.26	19.63	11.25	10.35	11.31	10.40	5.40	5.40	5.41	5.41
	Non-Markov	1.21	2.14	1.94	1.34	23.91	20.11	19.35	13.22	12.29	36.17	34.83	7.42	28.31	12.10	39.84
Non-Markov process with exponential distribution	Markov	1.28	1.88	1.95	1.23	23.34	20.22	20.06	25.10	23.35	1251.50	1273.01	107.85	4.20	893.30	3582.44
	Non-Markov	1.28	1.88	1.95	1.23	23.58	20.46	19.97	0.80	1.62	33.86	43.10	1.33	10.97	20.63	171.59
Markov process with Weibull distribution	Markov	8.95	8.15	14.28	6.21	116.55	39.88	87.35	84.49	135.21	84.12	134.50	96.33	96.33	96.31	96.31
	Non-Markov	8.96	8.15	14.29	6.21	116.60	39.87	87.31	82.96	134.74	79.67	132.29	98.77	97.29	100.82	86.86
Non-Markov process with Weibull distribution	Markov	8.82	8.47	14.57	6.47	115.56	40.39	87.90	66.66	86.43	1825.49	1695.88	103.89	329.20	1103.26	888.71
	Non-Markov	8.82	8.47	14.58	6.47	115.49	40.39	87.93	13.84	16.50	67.41	113.79	0.86	16.36	96.31	136.29
Scenarios	Models	Width of 95% CI														
Markov process with exponential distribution	Markov	3.73	4.52	4.52	3.73	16.85	15.52	15.40	10.87	10.87	10.88	10.87	6.33	6.33	6.33	6.32
	Non-Markov	3.73	4.52	4.52	3.73	16.89	15.52	15.45	12.02	12.11	22.51	22.56	8.55	19.22	12.11	23.74
Non-Markov process with exponential distribution	Markov	3.72	4.50	4.50	3.71	17.22	15.58	15.62	3.86	4.92	3.87	4.91	4.35	4.35	4.35	4.35
	Non-Markov	3.71	4.50	4.50	3.71	17.24	15.56	15.59	3.38	4.73	21.86	23.59	4.31	12.39	15.32	32.44
Markov process with Weibull distribution	Markov	4.71	5.36	5.38	4.68	10.86	10.32	10.15	19.04	19.45	19.05	19.43	5.53	5.53	5.53	5.53
	Non-Markov	4.71	5.36	5.38	4.68	10.86	10.31	10.15	19.89	20.29	24.04	24.80	7.36	16.75	11.19	19.25
Non-Markov process with Weibull distribution	Markov	4.71	5.35	5.37	4.67	10.84	10.32	10.13	2.84	3.53	2.84	3.53	3.02	3.02	3.02	3.02
	Non-Markov	4.71	5.35	5.37	4.67	10.85	10.31	10.13	1.69	2.52	21.38	22.11	2.30	14.14	23.16	18.27
Scenarios	Models	Coverage rate (%)														
Markov process with exponential distribution	Markov	91	88	90	91	92	94	93	90	91	89	91	80	80	80	80
	Non-Markov	91	88	90	91	92	94	94	90	91	93	94	87	91	91	92
Non-Markov process with exponential distribution	Markov	91	90	90	91	93	93	94	2	15	0	0	0	70	0	0
	Non-Markov	91	91	89	91	93	93	94	94	93	93	92	93	93	89	80
Markov process with Weibull distribution	Markov	49	63	42	60	21	48	24	80	65	80	65	14	14	14	14
	Non-Markov	49	63	43	60	21	48	24	82	66	84	70	18	60	35	69
Non-Markov process with Weibull distribution	Markov	49	62	41	59	21	48	23	0	15	1	0	24	0	0	0
	Non-Markov	49	62	42	59	22	49	23	87	58	84	69	73	92	70	41

TR: transition rates. $TR_{23|0}$: Transition rate from state 2 to state 3 with a history of state 0. $TR_{34|0_1_2}$: Transition rate from state 3 to state 4 with a history of states 0, 1, and 2.

Figure 4. Age-specific transition rates (cases per 100 persons) from age 45 to 94 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

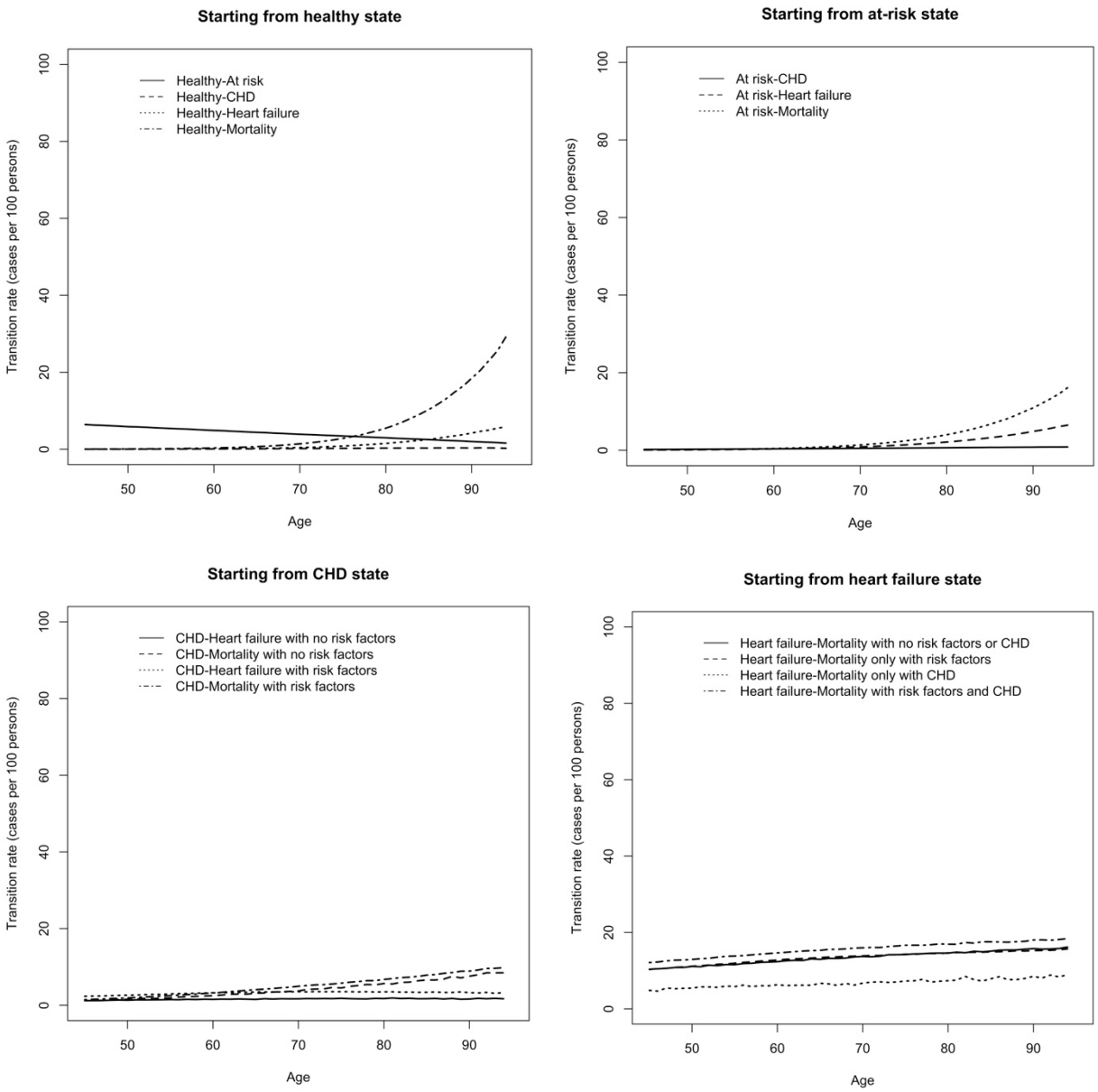


Figure 5. Transition probabilities (%) starting from age 45 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

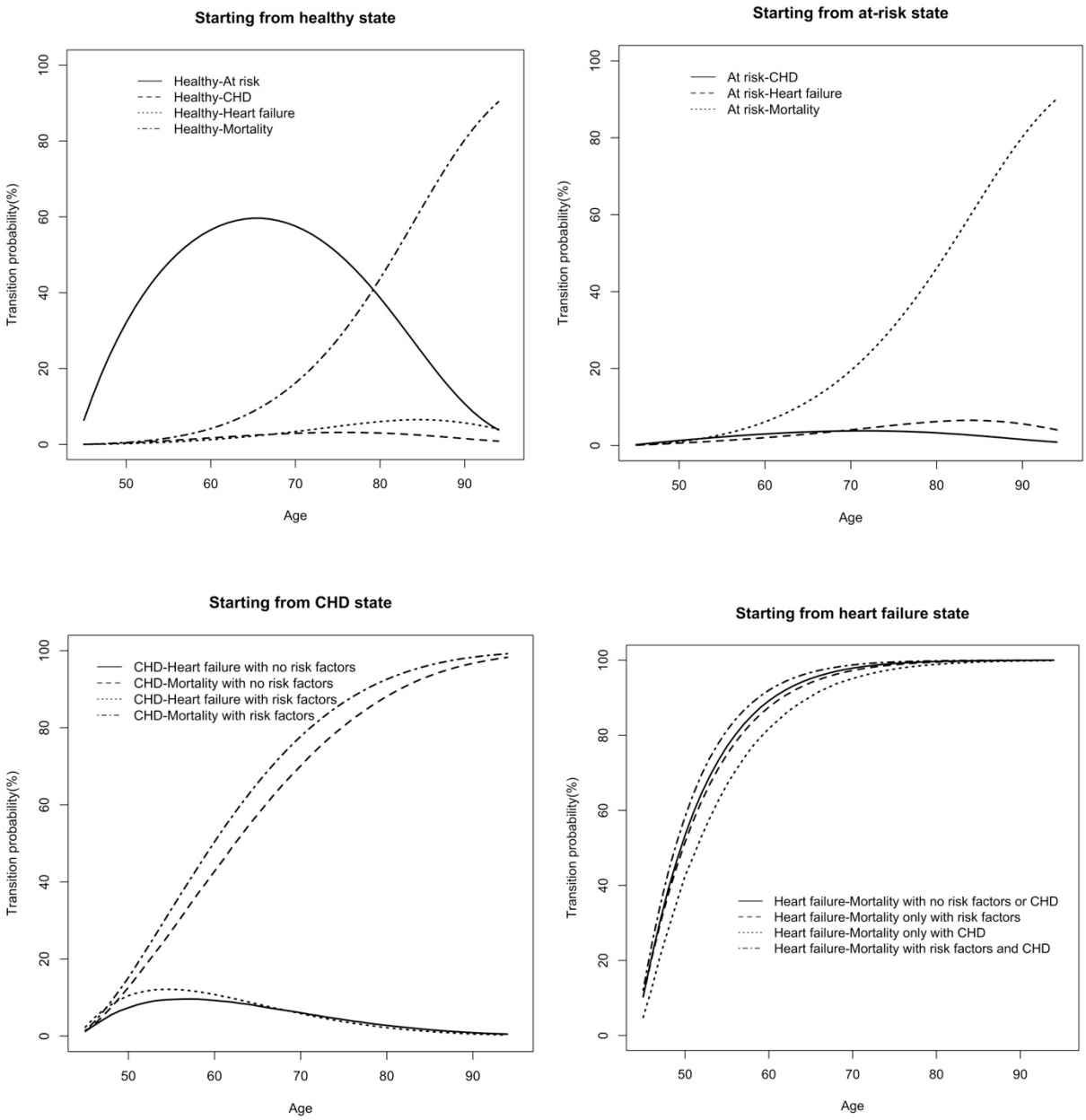


Figure 6. State occupational probabilities (%) starting from age 45 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

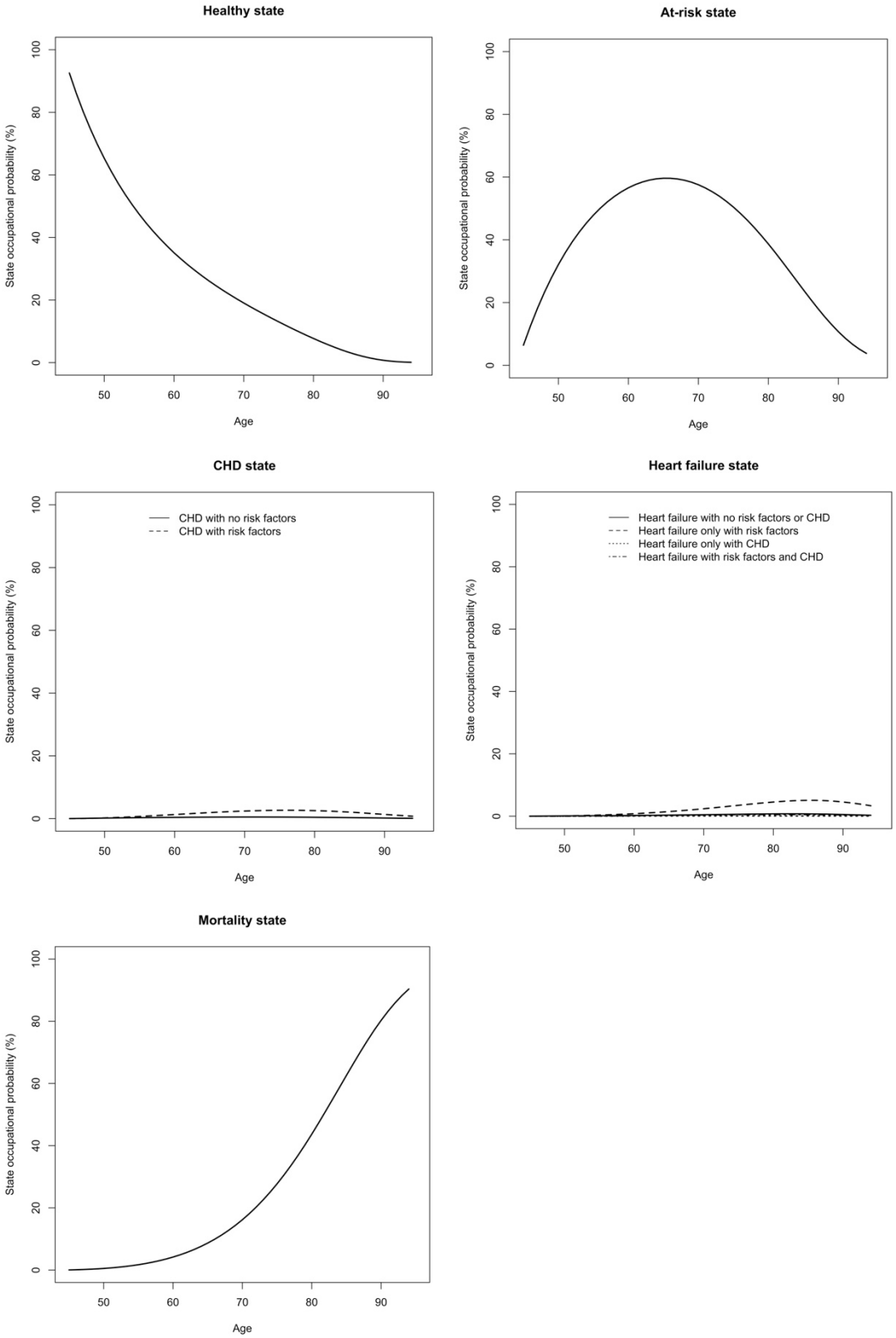


Table 3. Transition rates (cases per 100 persons) at age 70 and 90 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 70 years	Age 90 years
Transition from healthy state		
Healthy → at-risk	4.21 (3.99, 4.41)	2.39 (2.12, 2.66)
Healthy → CHD	0.21 (0.17, 0.26)	0.75 (0.34, 1.15)
Healthy → heart failure	0.45 (0.38, 0.52)	5.13 (3.78, 6.52)
Healthy → mortality	1.34 (1.22, 1.47)	19.12 (16.39, 21.63)
Transition from at-risk state		
At-risk → CHD	0.51 (0.48, 0.54)	0.91 (0.78, 1.03)
At-risk → heart failure	0.83 (0.80, 0.88)	4.85 (4.46, 5.27)
At-risk → mortality	1.28 (1.23, 1.33)	10.58 (10.00, 11.25)
Transition from CHD state with no risk factors		
CHD → heart failure	2.81 (1.69, 3.94)	3.42 (1.58, 5.24)
CHD → mortality	5.25 (3.60, 6.84)	12.02 (7.14, 16.51)
Transition from CHD state with risk factors		
CHD → heart failure	3.99 (3.51, 4.43)	4.87 (3.49, 6.41)
CHD → mortality	5.27 (4.72, 5.77)	11.96 (8.87, 15.26)
Transition from heart failure state with no risk factors or CHD		
Heart failure → mortality	15.80 (13.59, 17.91)	19.33 (15.61, 23.06)
Transition from heart failure state only with risk factors		
Heart failure → mortality	14.60 (13.78, 15.45)	17.82 (15.12, 20.32)
Transition from heart failure state only with CHD		
Heart failure → mortality	13.43 (6.21, 20.91)	16.96 (7.96, 25.24)
Transition from heart failure state with risk factors and CHD		
Heart failure → mortality	17.65 (15.91, 19.61)	21.54 (17.67, 25.02)

Table 4. Transition probabilities (%) from age 45 years to age 70 and 90 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 70 years	Age 90 years
Transition from healthy state		
Healthy → Healthy	20.68 (20.35, 21.00)	1.17 (1.09, 1.25)
Healthy → at risk	58.67 (58.37, 58.96)	13.28 (13.05, 13.52)
Healthy → CHD	2.91 (2.85, 2.98)	1.85 (1.72, 1.97)
Healthy → heart failure	3.20 (3.13, 3.26)	6.22 (5.97, 6.50)
Healthy → mortality	14.54 (14.40, 14.69)	77.48 (77.11, 77.82)
Transition from at-risk state		
At risk → At risk	74.50 (74.32, 74.71)	14.34 (14.08, 14.60)
At risk → CHD	3.84 (3.75, 3.92)	1.89 (1.74, 2.04)
At risk → heart failure	3.91 (3.83, 3.98)	6.17 (5.88, 6.48)
At risk → mortality	17.76 (17.59, 17.90)	77.61 (77.21, 77.97)
Transition from CHD state with no risk factors		
Remain in CHD state	22.39 (20.75, 24.06)	1.90 (1.58, 2.28)
CHD → heart failure	7.69 (6.42, 9.21)	1.31 (1.02, 1.71)
CHD → mortality	69.90 (67.81, 72.06)	96.78 (96.20, 97.27)
Transition from CHD state with risk factors		
Remain in CHD state	17.10 (16.52, 17.71)	1.10 (0.97, 1.23)
CHD → heart failure	6.59 (6.30, 6.93)	0.73 (0.66, 0.81)
CHD → mortality	76.30 (75.62, 76.99)	98.17 (98.01, 98.32)
Transition from heart failure state with no risk factors		
Remain in heart failure state	2.21 (1.96, 2.52)	0.05 (0.04, 0.06)
Heart failure → mortality	97.79 (97.48, 98.04)	99.95 (99.94, 99.96)
Transition from heart failure state only with risk factors		
Remain in heart failure state	3.08 (2.95, 3.23)	0.09 (0.08, 0.10)
Heart failure → mortality	96.92 (96.77, 97.05)	99.91 (99.90, 99.92)
Transition from heart failure state only with CHD		
Remain in heart failure state	3.96 (2.79, 5.59)	0.14 (0.09, 0.25)
Heart failure → mortality	96.04 (94.41, 97.21)	99.86 (99.75, 99.91)
Transition from heart failure state with risk factors and CHD		
Remain in heart failure state	1.40 (1.27, 1.53)	0.02 (0.02, 0.02)
Heart failure → mortality	98.60 (98.47, 98.73)	99.98 (99.98, 99.98)

Table 5. State occupational probabilities (%) from age 45 years to age 70 and 90 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 70 years	Age 90 years
Healthy	20.68 (20.37, 21.01)	1.17 (1.09, 1.26)
At-risk	58.67 (58.38, 58.94)	13.28 (13.06, 13.52)
CHD with no risk factors	0.54 (0.50, 0.58)	0.24 (0.21, 0.28)
CHD with risk factors	2.38 (2.33, 2.43)	1.61 (1.49, 1.74)
Heart failure with no risk factors or CHD	0.48 (0.44, 0.51)	0.65 (0.59, 0.72)
Heart failure only with risk factors	2.24 (2.19, 2.29)	5.00 (4.76, 5.25)
Heart failure only with CHD	0.09 (0.07, 0.11)	0.08 (0.06, 0.10)
Heart failure with risk factors and CHD	0.39 (0.37, 0.41)	0.49 (0.44, 0.54)
Mortality	14.54 (14.39, 14.70)	77.47 (77.13, 77.84)

Table S1. Parameters used for simulation study using the 'crisk.sim' package in R.

Status	Markov process with exponential distribution		Markov process with Weibull distribution		Status	Non-Markov process with exponential distribution		Non-Markov process with Weibull distribution	
	Ancillary parameters	beta0	Ancillary parameters	beta0		Ancillary parameters	beta0	Ancillary parameters	beta0
Dataset 1					**Dataset 1**				
S_0 to S_1	1	-log(0.2)	1.5	-log(0.2)	S_0 to S_1	1	-log(0.2)	1.5	-log(0.2)
S_0 to S_2	1	-log(0.3)	1.3	-log(0.3)	S_0 to S_2	1	-log(0.3)	1.3	-log(0.3)
S_0 to S_3	1	-log(0.3)	1.5	-log(0.3)	S_0 to S_3	1	-log(0.3)	1.5	-log(0.3)
S_0 to S_4	1	-log(0.2)	1.4	-log(0.2)	S_0 to S_4	1	-log(0.2)	1.4	-log(0.2)
Dataset 2					**Dataset 2**				
S_1 to S_2	1	-log(0.4)	2	-log(0.4)	S_1 to S_2	1	-log(0.4)	2	-log(0.4)
S_1 to S_3	1	-log(0.3)	1.3	-log(0.3)	S_1 to S_3	1	-log(0.3)	1.3	-log(0.3)
S_1 to S_4	1	-log(0.3)	2	-log(0.3)	S_1 to S_4	1	-log(0.3)	2	-log(0.3)
Dataset 3					**Dataset 3.1**				
S_2 to S_3	1	-log(0.5)	0.9	-log(0.3)	$S_{2\|0}$ to S_3	1	-log(0.1)	0.3	-log(0.1)
S_2 to S_4	1	-log(0.5)	0.9	-log(0.3)	$S_{2\|0}$ to S_4	1	-log(0.2)	0.5	-log(0.2)
					Dataset 3.2				
					$S_{2\|0_1}$ to S_3	1	-log(0.5)	1.2	-log(0.5)
					$S_{2\|0_1}$ to S_4	1	-log(0.6)	1.4	-log(0.6)
Dataset 4									
S_3 to S_4	1	-log(0.5)	1.8	-log(0.4)	**Dataset 4.1**. $S_{3\|0}$ to S_4	1	-log(0.2)	0.5	-log(0.2)
					Dataset 4.2. $S_{3\|0_1}$ to S_4	1	-log(0.3)	1.2	-log(0.3)
					Dataset 4.3. $S_{3\|0_2}$ to S_4	1	-log(0.6)	2	-log(0.6)
					Dataset 4.4. $S_{3\|0_1_2}$ to S_4	1	-log(0.9)	3	-log(0.9)