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

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, investigation of disease etiology has largely focused on one single endpoint, and progression of chronic disease as a multi-state process is understudied, representing a knowledge gap. Most of existing multi-state regression models require Markov assumption and are unsuitable to estimate progression of chronic diseases that is largely non-memoryless. We propose a new non-Markov framework that allows past states to affect transition rates of current states. The key innovation is that we convert a non-Markov to Markov process by dividing disease states into substates through conditioning on past disease history. Specifically, we apply cause-specific Cox models (CSC) including past states as covariates to obtain transition rates (TR) of substates, which were used to obtain transition probabilities (TP) and state occupational probabilities (SOP) of substates. We applied our model to describe progression of coronary heart disease (CHD) in the ARIC study, where CHD was modeled in healthy, in risk, CHD, heart failure, and mortality states. We presented transition rates, transition probabilities, and state occupational probabilities between states from age 45 to 95 years. In summary, the significance of our framework lies in that transition parameters between disease substates may shed light on new mechanistic insight of chronic disease and may provide more accurate description of non-Markov process than Markov regression models. Our method has potential of wide application in chronic disease epidemiology.

1. INTRODUCTION

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

Most of the existing methods describing progression of chronic disease are Markov models, which assume that future state depends only on the present state and not on past history. Transition probabilities of Markov process can be parametrically estimated 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 transitions rate, Markov regression models have been proposed that allow to include covariates into the model. For example, by dividing data into subsets according to disease states and changing data structure in each subset from wide to long, Cox Markov regression applies Cox model within each subset to estimate transition rates between states,[1] and the discrete-time transition rates estimated from the model are plugged into A-J estimators to obtain transition probabilities.[8] However, Markov models are unsuitable to model progression of chronic diseases, which is largely non-memoryless, and this signifies the urgent need of non-Markov methods to advance the research in estimation of disease progression.

Non-Markov models are a largely unexplored field, with only several non-parametric methods proposed.[9-15] Although A-J estimator can consistently estimates transition probabilities

from time 0 to t regardless of Markov assumption,[11, 16, 17] it is systematically biased for any two random time points in non-Markov process.[13] The landmark Aalen-Johansen (LMAJ) estimator accounts for non-Markov assumption and estimates transition probabilities using land-marking by dividing participants into subgroups according to the state occupation probability at a certain time point.[9] However, the subsampling of participants lowers the power of estimation and it is a non-parametrical method which may be difficult to incorporate covariates.[18, 19]

In this paper, we propose a new non-Markov regression model that allows past states to affect transition rates of current states. Compared to Markov regression models, the key innovation is that by conditioning on past disease history, we divide disease states into substates to convert non-Markov to Markov process. Specifically, we apply cause-specific Cox models (CSC) including past states as covariates to obtain transition probability (TP), transition rates (TR), and state occupational probability (SOP) between substates.

2. METHODS

A multi-state non-Markov regression method. While our method can be used for chronic disease of any given states, we illustrate our method assuming five states S_0-S_4 (**Fig 1a**). By conditioning on past states, a disease state can be divided into substates, and non-Markov process can be converted to Markov process (**Fig 1b**). Suppose we have survival data with disease states ascertained (**Fig 2a**). We subset the data into subsets 1-4 by disease states (**Fig 2b**) and change the data structure from wide to long by age or calendar time within each subset (**Fig 2c**).

2.1. Estimate transition rate, Q(t), is defined as rate of state transition at time t. As the transition of chronic disease from one to multiple states is a competing risk question, CSC models are used to estimate transition rates between substates. As an extension of Cox model,

CSC incorporates cause-specific hazards for the competing events and assumes different associations of each exposure with each specific outcome.[20] The risks of developing the competing outcomes over time can be estimated using the cumulative incidence function of CSC, which has been implemented in 'predict' command of 'riskRegression' package in R.[21] To estimate transition rates, by changing data subsets into long format and modeling interval as a categorical variable, the predicted risk by the end of each interval is approximate to hazard in discrete time, which can be estimated by modeling and using 'predict' command of 'riskRegression' package. Particularly, four CSC models are fitted to data subsets 1-4 (**Models 1-4**). The influence of past history of disease on transition rates can be accounted for in models 3 and 4 in two ways. First, past disease history (as well as their interaction terms with other covariates) can be included as covariates in the model. Second, the dataset can be stratified by past disease history with transition rates estimated within each strata.

Model 1 in subset 1 starting from S_0 state: $h_c(t,X)=h_{0c}(t)\exp(\boldsymbol{\beta}_c * \text{interval} + \sum_j \boldsymbol{\beta}_{jc}x_j)$, 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: $h_c(t,X)=h_{0c}(t)\exp(\boldsymbol{\beta}_c * \text{interval} + \sum_j \boldsymbol{\beta}_{jc}x_j)$, where c= rate of developing S_2, S_3, or S_4 states

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

Transition rates can also be estimated using stratified analysis with past S_1 state as strata.

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

Transition rates can also be estimated using stratified analysis with past S_1 and S_2 states as strata.

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 proportion of participants occupied in each state at t and can be expressed as functions of transition rates. As $\mathbf{P}'(t) = \mathbf{P}(t) * \mathbf{Q}(t)$, $\mathbf{P}(t+1) = \mathbf{P}(t) + \mathbf{P}'(t) = \mathbf{P}(t) + \mathbf{P}(t) * \mathbf{Q}(t) = \mathbf{P}(t) * (I_9 + \mathbf{Q}(t)) = \left(\mathbf{P}(t-1) + \mathbf{P}'(t-1)\right) * \left(I_9 + \mathbf{Q}(t)\right) = \mathbf{P}(t-1) * (I_9 + \mathbf{Q}(t-1)) * (I_9 + \mathbf{Q}(t)) = \mathbf{P}(t=0) * \prod_{t=0}^{t=k}(I_9 + \mathbf{Q}(t))$, where $\mathbf{P}(t) = \left(p_0(t), p_1(t), p_{2|0}(t), p_{2|0_1}(t), p_{3|0}(t), p_{3|0_1}(t), p_{3|0_2}(t), p_{3|0_1_2}(t), p_4(t)\right)$, $\mathbf{Q}(t)$ is the transition rate matrix, and I_9 is an identity matrix.

2.3. Transition probabilities, TP (k_1, k_2), is defined as probability of state transition over a period, which can be estimated using two methods. First, for transition starting from state m, the transition probability from time k_1 to k_2 can be estimated as $\mathbf{TP_m}(k_1, k_2) = \prod_{t=k1}^{t=k2}(I + \mathbf{Q_m}(t))$, where $\mathbf{Q_m}(t)$ is the transition rate from state m to following states. Second, CSC models can be applied to datasets in wide format in **Fig 2b**, with intervals removed as covariates from models 1-4. The predicted risks are transition probabilities from starting time to the end of each time interval. Each method of calculating TP has its advantages. With $\mathbf{Q_m}(t)$ at each time known, the first method is more flexible with starting time of transition. The second method may be more accurate, as transition probabilities are estimated in continuous time rather than discrete time.

2.4. Use bootstrap to obtain 95% CIs. We use bootstrap to repeatedly draw transition rates from their corresponding sampling distributions to estimate the sampling variability of the dynamic parameters. In each iteration, we randomly sample transition rates from the outputs

from CSC models. We repeat the sampling process for 1000 times, and the resulting 1000 calculated values reflect the variability of the parameters. The generated distributions for dynamic parameters are summarized with median values and 95% CIs defined by the 2.5th and 97.5th percentiles of the generated values.

3. APPLICATION STUDY

3.1. 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.[22] 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.[23] For each event, the month and year of diagnosis were recorded as diagnosis date. Heart failure was ascertained by surveillance calls or clinic visits, and has 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.[24] 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.

3.2. Methods. We modeled coronary heart disease (CHD) progression in five states: Healthy, in risk, CHD, heart failure, and mortality. In-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.[25] 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).[26] Diabetes was defined as fasting glucose ≥7.7mmol/L.[27] We calculated time of incidence of in-risk state as the earliest time that any of the risk factors was documented.

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

We applied our non-Markov framework to examine dynamics of CHD progression. Briefly, we divided the longitudinal data into four subsets, "individuals start from healthy state to incidence of hypertension, CHD, heart failure, or mortality, whichever outcome came first", "individuals start from hypertension 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 in long format, we applied CSC models by including age and past disease history (models 3 and 4 only) as covariates into the model, where age was modeled flexibly in cubic terms.

3.3. Results. Our study included 15027 participants who did not develop CHD or heart failure at baseline. During a median of 27 years of follow up, our study documented 13043 in-risk participants, 2565 incident cases of CHD, 3283 incident cases of heart failure, and 7677 incident

cases of mortality. The transition rates from healthy, in-risk or CHD states to subsequent states were low at mid-age and gradually increased with age (**Figure 4**). There was a drastically increase in rate of transition from healthy or in-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 60 and 75 years in **Table 1**.

We estimated transition probability from each state to the following states starting from 45 years old (**Figure 5**). A person started from a healthy state was very likely to develop in-risk state over follow up. A person started from in-risk state was more likely to develop CHD than heart failure or mortality before 70 years, however, the risk of developing heart failure or mortality from in-risk state increased sharply after 70 years and exceeded the risk of developing CHD. 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 60 and 75 years in **Table 2**.

We estimated state occupational probability for participants starting from healthy states at age 45 years. The proportion of participants in healthy state decreased over time and the proportion in mortality state increased over time (**Figure 6**). The proportion of in-risk participants first increased and then decreased, which was due to that these participants moved on to 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, in-risk, CHD, heart failure, and mortality states at age 75 years were 1.17, 13.28, 1.85, 5.64, and 78.04 %, respectively (**Table 3**). Most of the participants in CHD or heart failure states had a history of risk factors: 87% of CHD participants had risk factors and 90% of heart failure participants had risk factors or CHD history.

4. DISCUSSION

We proposed a new non-Markov regression model that relaxes Markov assumption by allowing past states to affect current states. The innovation lies in that a non-Markov process is converted to Markov process by conditioning on past states and estimating transition parameters between substates. The application study showed that our method may initiate a new direction of investigating disease progression in cardiovascular epidemiology.

As the leading cause of mortality in the U.S.,[28] 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,[29] including smoking,[30] body mass index (BMI),[31-33] diet quality,[34-36] and physical activity.[37, 38] However, these studies largely focused on one single endpoint, such as incidences of hypertension, CHD, or heart failure.[39-44] A paucity of studies examined CHD progression in small sample size using Markov models, leading to imprecise estimation and limited generalizability of the findings.[45-48] The implication of our application study lies in two aspects. First, the estimated age-specific transition rates indicate disease mechanism and can guide disease state-specific precision strategies targeting participants in 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 early states in prevention of 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, the 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.[49, 50] 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 more accurate estimation of disease burden and can facilitate medical resources planning for real-time. In addressing the significance of our study, we acknowledge that we leveraged the unique advantages of ARIC data for our model application, including the high follow-up rates, CHD endpoints ascertained by medical records, the accumulation of substantial data to capture the long-term CHD course, and the accessible of the data through an NHLBI data repository BioLINCC.

In summary, we developed a multi-state non-Markov framework that allows past states to affect transition rates of current states. The transition parameters between substates may enhance our understanding of new mechanisms of disease progression and stimulate new approaches of disease prevention. Our framework is highly suitable to estimate progression of chronic diseases that are non-memoryless and has potential wide application in chronic disease epidemiology.

ACKNOWLEDGMENTS The project described was supported by National Institutes of Health and National Center for Advancing Translational Sciences (NCATS), through Grant Award Number R21 HG012365 and 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

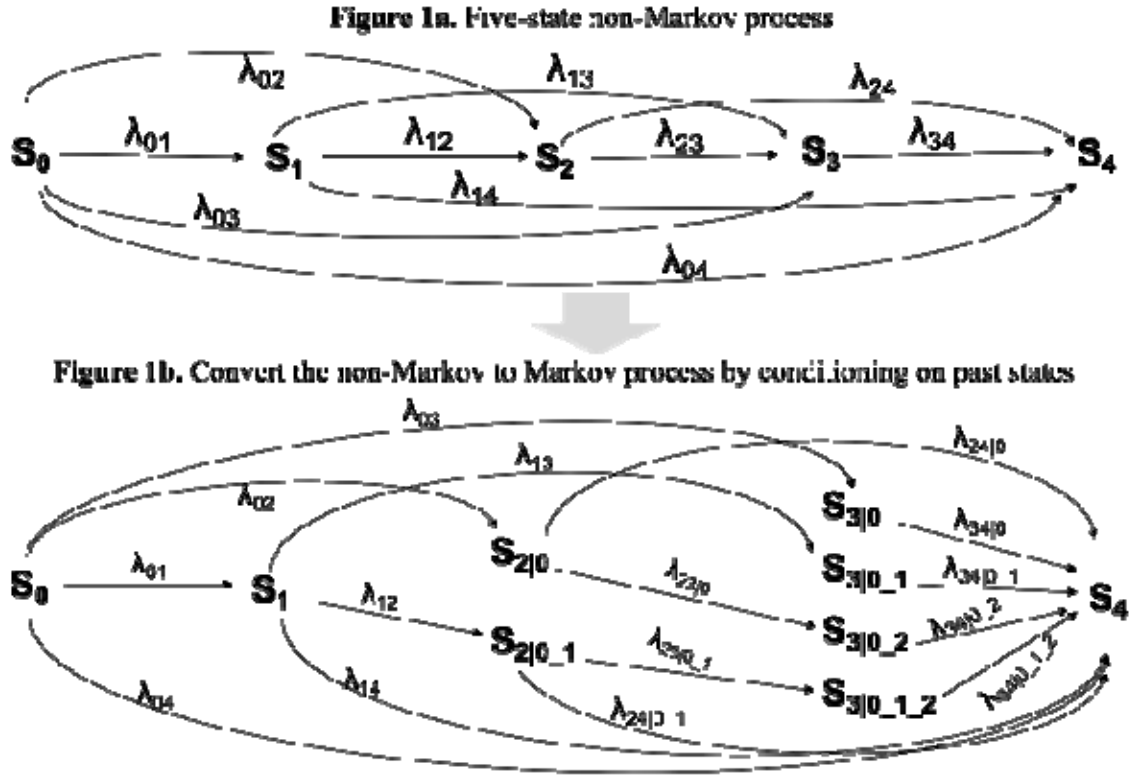
Biomed. 2010;99(3):261-74. Epub 2010/03/17. doi: 10.1016/j.cmpb.2010.01.001. PubMed PMID: 20227129.
9. Putter H, Spitoni C. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. Stat Methods Med Res. 2018;27(7):2081-92. Epub 2018/05/31. doi: 10.1177/0962280216674497. PubMed PMID: 29846146.
10. 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.
11. Pepe MS. Inference for Events With Dependent Risks in Multiple Endpoint Studies. Journal of the American Statistical Association. 1991;86(415):770-8.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.

23. Rose GA. Cardiovascular survey methods. Geneva Albany, N.Y.: World Health Organization; WHO Publications Centre distributor1982.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.

43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.

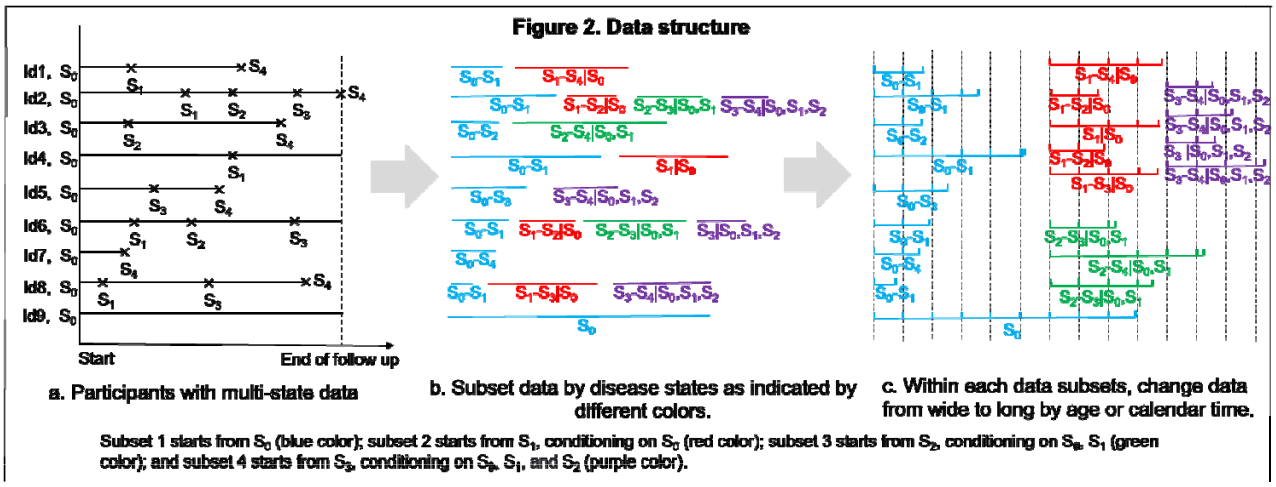


Figure 3. Transition rate matrix Q(t).

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

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

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.

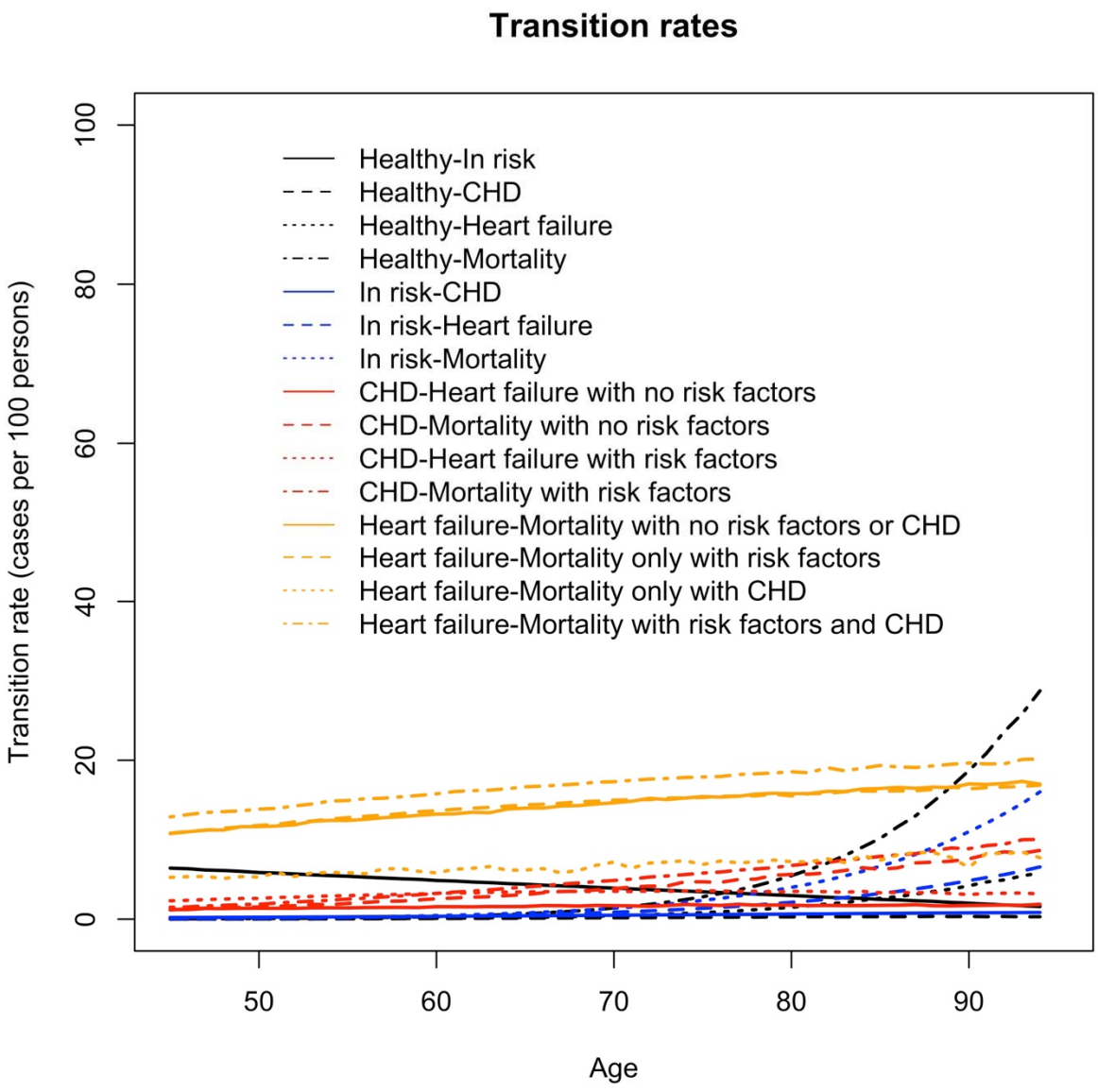


Figure 5. Transition probabilities (%) starting from age 45 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

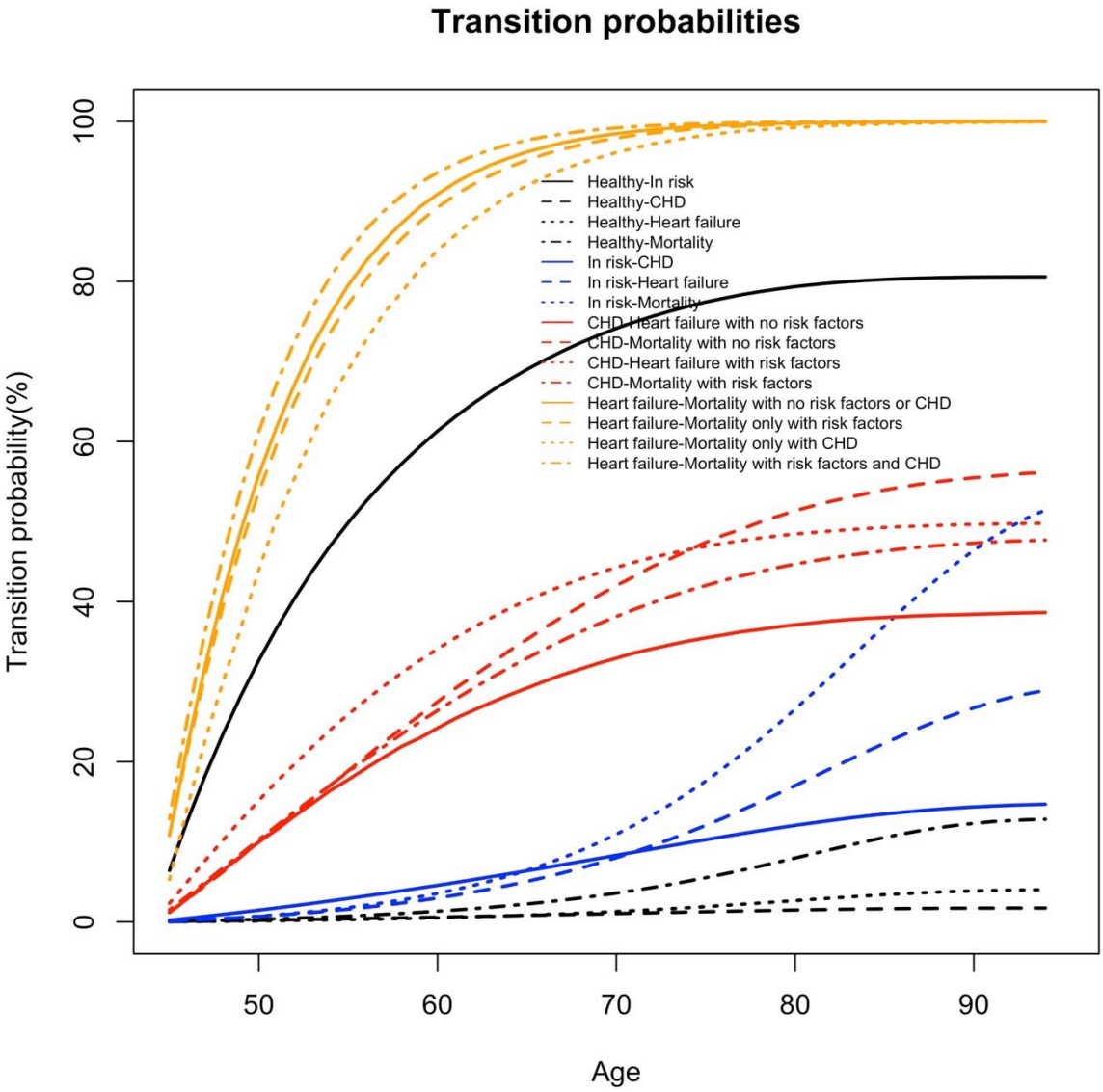


Figure 6. State occupational probabilities (%) starting from age 45 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

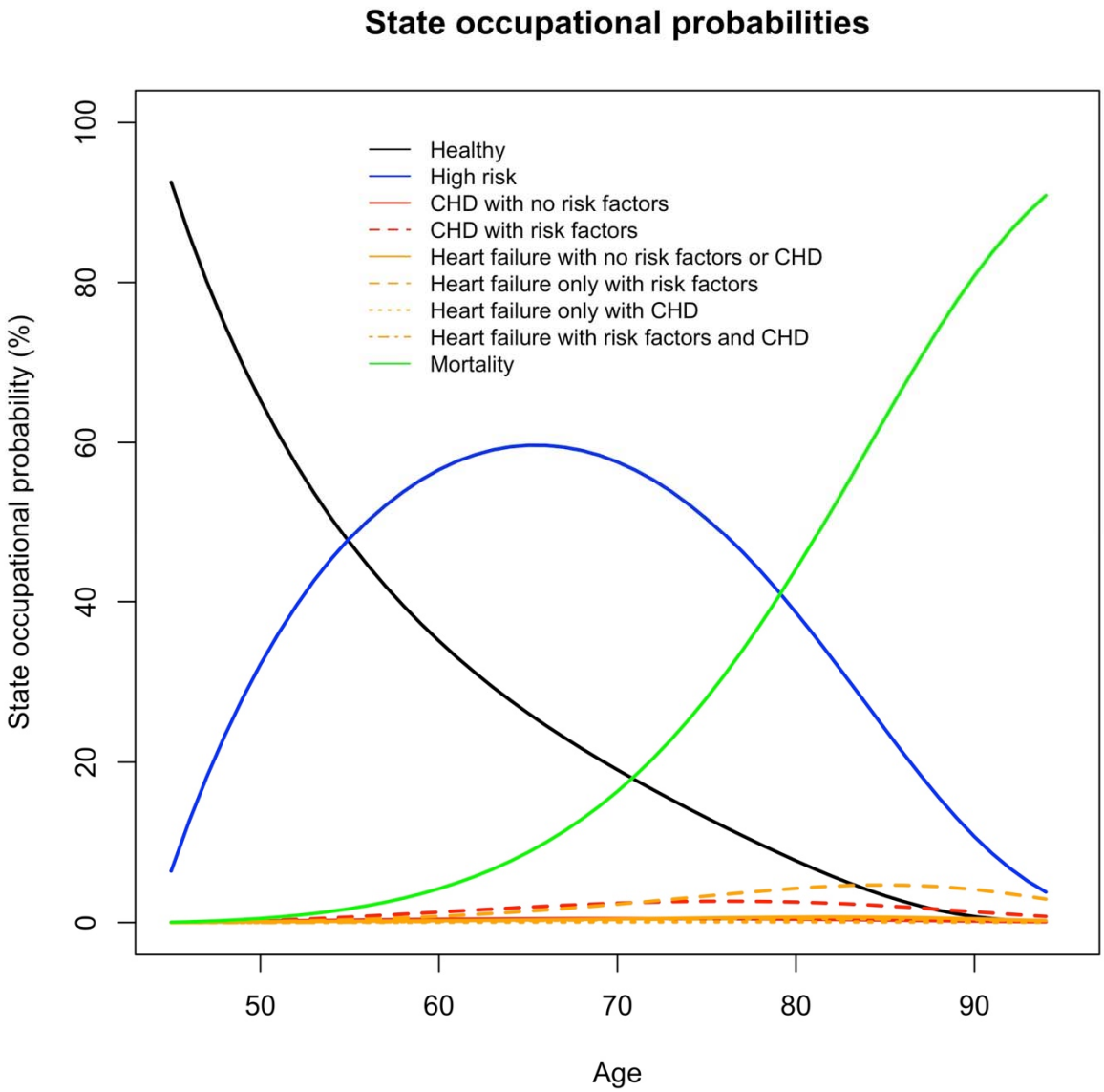


Table 1. Transition rates (cases per 100 persons) at age 60 and 75 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 60 years	Age 75 years
Transition from healthy state		
Healthy to in-risk	4.21 (4.01, 4.42)	2.38 (2.12, 2.68)
Healthy to CHD	0.21 (0.17, 0.26)	0.75 (0.33, 1.15)
Healthy to heart failure	0.45 (0.38, 0.52)	5.16 (3.82, 6.48)
Healthy to mortality	1.34 (1.22, 1.46)	19.05 (16.54, 21.56)
Transition from in-risk state		
In-risk to CHD	0.51 (0.48, 0.54)	0.91 (0.78, 1.04)
In-risk to heart failure	0.83 (0.79, 0.87)	4.87 (4.43, 5.27)
In-risk to mortality	1.28 (1.23, 1.33)	10.58 (9.91, 11.28)
Transition from CHD state with no risk factors		
CHD to heart failure	2.83 (1.68, 3.91)	3.47 (1.63, 5.14)
CHD to mortality	5.24 (3.64, 6.82)	12.14 (7.78, 16.79)
Transition from CHD state with risk factors		
CHD to heart failure	3.97 (3.50, 4.44)	4.84 (3.31, 6.38)
CHD to mortality	5.24 (4.73, 5.75)	11.97 (8.44, 15.18)
Transition from heart failure state with no risk factors or CHD		
Heart failure to mortality	17.34 (14.57, 19.98)	21.42 (16.56, 25.97)
Transition from heart failure state only with risk factors		
Heart failure to mortality	15.80 (14.78, 16.73)	19.69 (16.53, 22.74)
Transition from heart failure state only with CHD		
Heart failure to mortality	14.68 (6.69, 23.48)	18.12 (8.33, 28.55)
Transition from heart failure state with risk factors and CHD		
Heart failure to mortality	19.41 (17.24, 21.61)	23.95 (19.14, 28.55)

Table 2. Transition probabilities (%) from age 45 years to age 60 and 75 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 60 years	Age 75 years
Transition from healthy state		
Remain in healthy state	20.66 (20.36, 20.98)	1.17 (1.09, 1.26)
Healthy to in-risk	73.65 (73.28, 74.02)	80.81 (80.51, 81.09)
Healthy to CHD	1.07 (0.99, 1.15)	1.79 (1.70, 1.89)
Healthy to heart failure	1.24 (1.18, 1.31)	3.93 (3.79, 4.05)
Healthy to mortality	3.37 (3.26, 3.47)	12.30 (12.05, 12.54)
Transition from in-risk state		
Remain in in-risk state	74.50 (74.32, 74.68)	14.34 (14.07, 14.59)
In-risk to CHD	8.06 (7.93, 8.19)	14.39 (14.22, 14.56)
In-risk to heart failure	7.45 (7.34, 7.56)	26.22 (25.98, 26.46)
In-risk to mortality	9.99 (9.87, 10.12)	45.06 (44.74, 45.35)
Transition from CHD state with no risk factors		
Remain in CHD state	22.32 (20.52, 24.15)	1.89 (1.58, 2.28)
CHD to heart failure	34.67 (32.36, 36.93)	40.66 (38.45, 42.82)
CHD to mortality	43.02 (40.95, 45.21)	57.44 (55.36, 59.59)
Transition from CHD state with risk factors		
Remain in CHD state	17.13 (16.55, 17.76)	1.11 (0.98, 1.24)
CHD to heart failure	44.66 (43.51, 45.80)	50.70 (49.68, 51.77)
CHD to mortality	38.21 (37.22, 39.13)	48.19 (47.11, 49.16)
Transition from heart failure state with no risk factors		
Remain in heart failure state	1.60 (1.39, 1.86)	0.02 (0.02, 0.03)
Heart failure to mortality	98.40 (98.14, 98.61)	99.98 (99.97, 99.98)
Transition from heart failure state only with risk factors		
Remain in heart failure state	2.35 (2.23, 2.48)	0.05 (0.04, 0.05)
Heart failure to mortality	97.65 (97.52, 97.77)	99.95 (99.95, 99.96)
Transition from heart failure state only with CHD		
Remain in heart failure state	3.15 (2.07, 4.72)	0.08 (0.04, 0.15)
Heart failure to mortality	96.85 (95.28, 97.93)	99.92 (99.85, 99.96)
Transition from heart failure state with risk factors and CHD		
Remain in heart failure state	0.92 (0.82, 1.03)	0.01 (0.01, 0.01)
Heart failure to mortality	99.08 (98.97, 99.18)	99.99 (99.99, 99.99)

Table 3. State occupational probabilities (%) from age 45 years to age 60 and 75 years estimated using the non-Markov framework in the Atherosclerosis Risk in Communities Study (ARIC) study.

	Age 60 years	Age 75 years
Healthy	20.66 (20.36, 20.98)	1.17 (1.09, 1.26)
In-risk	58.68 (58.36, 58.96)	13.28 (13.04, 13.51)
CHD with no risk factors	0.54 (0.50, 0.58)	0.24 (0.21, 0.27)
CHD with risk factors	2.38 (2.33, 2.43)	1.61 (1.50, 1.75)
Heart failure with no risk factors or CHD	0.45 (0.42, 0.48)	0.58 (0.52, 0.65)
Heart failure only with risk factors	2.15 (2.09, 2.19)	4.56 (4.32, 4.83)
Heart failure only with CHD	0.09 (0.07, 0.10)	0.07 (0.05, 0.09)
Heart failure with risk factors and CHD	0.37 (0.35, 0.39)	0.43 (0.38, 0.49)
Mortality	14.69 (14.54, 14.84)	78.04 (77.69, 78.41)