Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2021;7(3): 285-304

Original Article

Multistate Models for the Analysis of Time to Type II Chronic Diabetic Complications in Debre Markos Referral Hospital, Northwest Ethiopia

Muluye Getie Ayaneh*, Ashagrie Sharew Iyasu

Department of Statistics, Debre Markos University, Debre Markos, Ethiopia.

ARTICLE INFO

Received 10.05.2021
Revised 27.07.2021
Accepted 03.08.2021
Published 25.09.2021

Key words:
Diabetes mellitus;
Vascular complications;
Multistate models

ABSTRACT

Introduction: Diabetes is a chronic, non-communicable disease characterized by elevated blood glucose levels. The purpose of this study was to jointly model the transition of diabetic patients in a series of clinical states and to assess the relationship between each state and different patient characteristics.
Methods: A hospital-based retrospective study was conducted on 524 patients with type II diabetes, aged 18 years or older, who attended their medication between January 1, 2005, and December 31, 2017. Multistate models with different assumptions were considered to explore the effects of different prognostic factors on the transition intensity of type II diabetes mellitus patients.
Results: During a median follow-up time of 7.4 years (Inter-Quartile Range=4.01), 54.8% of diabetic patients developed either microvascular or macrovascular complications, and 10.5% of them experienced both micro- and macrocomplications, and 16.66% of diabetes patients died. The assumption Markov was assessed by using the likelihood ratio test showed that Markov assumption was not held just for the transition. The transition rate of patients from the macrovascular state to the death state was affected by the residence of the patients (P=0.05) and age at diagnosis (p=0.01). The transition rates of patients with microvascular complications to death were significantly affected by baseline triglyceride level (P<0.001), age at first diagnosis (P=0.01), baseline glucose level (P=0.03, and baseline serum creatinine level (P=0.04).
Conclusion: The semi-Markov model fitted the data well and could be used as a convenient model for the analysis of time to diabetes-related complications or death.

Introduction

Diabetes mellitus (DM) is a chronic, non-communicable disease characterized by elevated levels of blood glucose that occurs due to failure in secretion, action, or both of insulin.[1] It is commonly classified as type I (T1DM), type II (T2DM), or gestational diabetes, of which T2DM is the most common form.[2] T1DM, where the possible risk factors are autoimmunity, genetic, and environmental factors, accounts for 5% to 10% of all cases

*.Corresponding Author: muluyegb@gmail.com

Copyright © 2021 Tehran University of Medical Sciences. Published by Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Noncommercial uses of the work are permitted, provided the original work is properly cited.

of diabetes, whereas type II diabetes occurs when the body cannot produce enough insulin or cannot use insulin, which accounts for 90% to 95% of all diagnosed diabetes cases.[3] T2DM may be unrecognized until complications become evident as the disease develops gradually.

Both the number of new cases and the prevalence of diabetes have been steadily increasing over the past few decades.[4] According to the International Diabetes Federation report, nearly half a billion people live with diabetes worldwide and among these, almost (80%) of diabetes cases are found in low- and middle-income countries. This happens because of rapid urbanization, unhealthy diets, increasingly sedentary lifestyles, and inadequate resources to provide preventive or medical care for their populations.[5] This is a major cause of death in most countries. As a result, diabetic complications have a significant economic impact on countries, healthcare systems, and above all, individuals with diabetes and their families.[6-8]

Ethiopia is challenged by the growing magnitude of chronic non-communicable diseases such as diabetes mellitus.[9] According to the 2017 International Diabetes Federation report, the number of people living with diabetes in Ethiopia was nearly 2.6 million, with an estimated prevalence of 3.8%.[6] The same report estimated that Ethiopia would be the 9th highest country globally, with 14.1 million diabetic cases by 2045.

Diabetes is not only a serious condition but also a risk factor for both short- and long- term complications. Microvascular complications (MiVasC's) and macrovascular complications (MaVasC's) are associated with long-term diabetes-related complications that can result in morbidity and mortality.[2, 8, 10] Diabetic nephropathy, neuropathy, and retinopathy are the main types of MiVasC's.[11] Diabetes mellitus and its complications have increased over the years and are among the common reasons for inpatient admissions in Ethiopia.[12] It is the ninth leading killer in the country.[13] The prevalence of diabetes-related vascular complications has not been well identified and documented, although its impact continues to rise. In studying chronic diseases such as diabetic complications, clinical interest lies in both the final outcome (death) and the dynamics of the complications process. Multistate models are useful in describing the progression of diseases with several possible states over time.[14-18] These models provide an appealing framework to study the movement of disease progression across a certain number of states defined by specific disease conditions and their prognostic factors.[19, 20]

Multistate models were introduced in 1978 by Aalen and Johansen.[21] However, these models are not commonly used despite their potential applications. Although multistate models are widely applied in clinical studies and other fields, they are often overlooked by investigators in time-to-event analysis.[22] Compared with the Cox regression model with a single endpoint, multistate models provide a more detailed insight into the disease process.[23, 24] Survival models in the multistate framework provide methods to explore the effect of covariates on multiple responses, taking into account and evaluating the effects of intermediate events. Therefore, the purpose of this study was to explore how T2DM patients progress in different disease status, and to determine the prevalence and predictors of vascular complications as well as mortality status of type II diabetes mellitus (T2DM)

patients based on a retrospective cohort study of patients in Debre Markos Referral Hospital (DMRH).

Specifically, this research sought to answer the following questions:

What is the prevalence of microvascular, and macrovascular complications among patients with T2DM?

To what extent do the possible risk factors affect the transition intensity of T2DM to different clinical states?

Is the effect of prognostic factors the same across transitions?

Methods

Study Area and Period

The study was conducted at the Debre Markos Referral Hospital (DMRH) in the town of Debre Markos. DMRH is located in the East Gojjam, Amhara National Regional State, which is 3000 KM in the northwest of Addis Ababa, capital of Ethiopia. The data for this study were collected from January 1, 2019, to March 30, 2019.

Study Design, Population and Period

A hospital-based retrospective study design was used on T2DM patients. All T2DM patients attending out-patient follow-up clinic at DMRH were taken as the source population. The study population was T2DM patients whose age was 18 years and above who have been on follow-up at DMRH from January 1, 2005, to December 31, 2017.

All T2DM patients were retrospectively followed-up for any medical condition (such as developing vascular complications or death) from entry to the end of follow up. Once the individual diagnosed to have T2DM, the individual are required to visit T2DM clinics every month under normal conditions or absence of comorbidities. The T2DM patient was considered right censored when she/he did not experience the event of interest or lost to follow-up in the follow-up period.

Data Collection Tools and Procedures

A data abstraction sheet was developed based on information documented in the patients' medical records. To ensure the quality of the data, nurses working in the same hospital were trained on data collection procedures and data abstraction sheets. The members of the research team supervised the data collection process to verify the completeness of the data. The data were checked for completeness during data entry and cleaning processes. The information obtained from the patient medical cards was entered into Excel sheets and subsequently transferred for analysis into R software.

Study Variables

The primary outcome (state) was defined as the time taken to develop vascular complications (MiVasC or MaVasC) or death from the patients' first T2DM illness but free of any complications. Covariates such as patient sex, diabetes duration with T2DM, age at the onset of diabetes, residence (rural or urban), blood pressure (<140/90mmHg or ≥140/90 mmHg), family history of diabetes, including fasting serum lipids (HDL, LDL, and triglycerides), glycated hemoglobin percentage (HbA1c%), serum creatinine level and serum electrolytes (sodium, potassium, and chloride levels) were

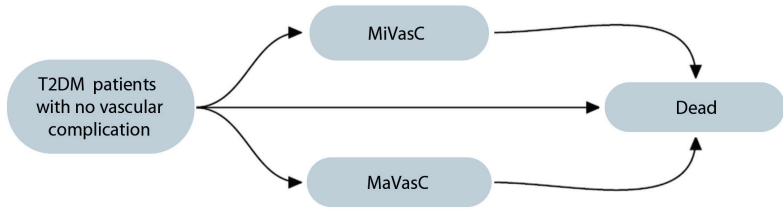

Figure 1. State structure that specifies states and possible transition for diabetic patients to their status of vascular complications or death

extracted. A patient was considered in the microvascular state if he/she developed at least one of the MiVasC's such as diabetic neuropathy, retinopathy, or nephropathy. Similarly, a patient was considered in the MaVasC state when a diabetic patient experienced an event of at least one MaVasC, such as peripheral artery disease, coronary artery disease, or stroke. The individual status of vascular complications was recorded only during clinical visits. Therefore, the exact date of an event is unknown (interval censoring). The right censoring time for each of the states (events) was measured for individuals who did not develop any of the complications during the follow-up period or were lost to follow-up.

Statistical Analyses

The baseline characteristics of the study participants were reported using descriptive statistics. A semi-parametric multistate model was applied to assess the relationship between the clinical conditions of patients and to determine the effects of covariates on individual states. The data were first entered into Microsoft Excel sheets and exported to R statistical software.[25] The mstate package in R software, which was developed by de Wreede et al.,[14] and the survival package written by Terry Therneau and Lumley[26] were used to carry out all analyses and generate graphs.

State Structure for Diabetic Complications

States were distinguished by vascular complication type and death. To formally describe the state structure of this study, we let "MiVasC", "MaVasC", and "died" denote T2DM patients who experienced microvascular compilation, macrovascular complications, and death, respectively. The state structure, which specifies the states and possible transitions, is shown in Figure 1. A patient who was complication-free was at risk of developing micro-or macro-vascular complications, or died with or without developing vascular complications at some time during the retrospective follow-up period.

Missing Data Analysis

The potential risk factors associated with T2DM were included in the transition intensity of the model as covariates. However, some values of the covariate were missing due to different reasons. Ignoring the presence of missing data might produce biased estimates that lead to a false conclusion when missing data are not missing completely at random (MCAR),[27] and even if the data are MCAR, dropping a large proportion of the data results in a substantial loss of information.[28]

To account for missing values in the covariates, we employed multiple imputations[28] to create

completed data sets with missing information filled in. This is an established and broadly accepted method to enhance data sets with missing information.[29] There are several methods that can be used to handle missing data that are based on the assumptions of the type and degree of 'missingness' in the dataset. R packages including Amelia II (uses Bootstrap EM),[30] Multivariate Imputation using Chained Equations (*mice*),[31] VIM,[32] and missForest[33] can be used to perform imputing missing data. For this analysis, the missForest R package was chosen as it was shown by Waljee et al.(2013) that it outperforms other well-known methods, and it is suitable for mixed-type data.[34] In missForest, missing values are imputed following the procedures and guidelines outlined in.[35]

Multistate Model for Diabetic Complication data

Multistate models are routinely used in research where a change in status over time is of interest. They have a wide range of applications, notably in chronic disease progression studies such as chronic vascular complications in diabetes patients where the study units reversibly or irreversibly move through a succession of events or states.[15, 24] It provides the opportunity to give a detailed description of the several states that individuals usually go through during a life course with type 2 diabetes.[36]
Let Mi(t) be the assumed state of individual *i*, where $i = 1, 2, \ldots, n$ at the instant $t = 1, 2, \ldots, r$ with respect to patient's T2DM complication status. Under the Markov assumption, the probability of an individual *i* moving from an *h* state to a *j* state in an interval δt given the patient's history (Hh−) is defined as:

$$P_{ihj}(\delta t) = P(M_i(t+\delta t = j \mid M_i(t) = h, H_h-) = P(M_i(t+\delta t) = j \mid M_i(t) = h) \quad (1)$$

If state *h* is not persistent, then $P_{hj}(s, t) \geq 0$ for and $P_{hj}(s, t) = 0$ for $h=j$. Otherwise, if state h is persistent, $P_{hj} = 0$. The transition intensity between states h and j is defined by equation (2)

$$\lambda_{ihj}(\delta t) = \lim_{\delta t \to 0} \frac{P(M_i(t+\delta t) = j \mid M_i(t) = h)}{\delta t} \quad (2)$$

Semi-parametric Multistate Model

The covariates in multistate models are often incorporated through transition intensity functions to explain differences among individuals during disease progression. One popular choice is the semi-parametric Cox regression model.[37] Under the proportional hazard assumption, the effect of a unit increase in a covariate is multiplicable with respect to the hazard rate.[38] According to de Wreede, Fiocco,[14] Andersen PK,[39] a time- homogeneous multistate model with proportional intensity having the transition-specific hazard, has the following form:

$$\lambda_{ij}(t \mid \mathbf{Z}) = \lambda_{ij0}(t) exp(\boldsymbol{\beta}_{ij}^{t} \mathbf{Z}), \; i \neq j \quad (3)$$

Here, $\lambda_{ij0}(t)$ is the baseline $i \rightarrow j$ transition intensity, which is common for all individuals and specifies how the intensity changes with time, while β_{ij} and *Z* are vectors of unknown regression coefficients and covariates, respectively, from the ith state to the j^{th} state transition. The model specified in Equation (3) is a parametric model when the baseline hazard specified by $\lambda_{ij0}(t)$ takes a functional

form. However, the covariates may be time-dependent, and the model becomes a nonhomogeneous Markov process.[40]

The model in Equation (3) assumes that the process is Markovian and is known as a Cox Markov model. Several limitations are encountered when ignoring the disease history. Therefore, an alternative approach is to use Cox semi-Markov models in which the future of the process depends on the duration of the patient in the current state rather than the current time.[29] Such models are also referred to as "clock reset" in which the time t in $\lambda_{hj}(t)$ is considered the time duration from the entrance of the patient to the state h.[17] The semi-Markov process considers that the sojourn time of transition from one state to another depends on the time already spent in the current state, which is the most important assumption in medical data.[41] The semi-Markov model for the progression of diabetic complications is specified in Equation.[4]

$$\lambda_{ij}(t-T_{ij}\mid \mathbf{Z}) = \lambda_{ij0}\left(t-T_{ij}\right)exp(\boldsymbol{\beta}_{ij}^{t}\mathbf{Z});\ i=1,2;\ j=2,3\ \ \&\ i<j \tag{4}$$

where T_{ij} is the sojourn time spent in the i^{th} state and $1 \le i < j \le 4$.[40]

Estimation Techniques

Under the Markov assumption, the nonparametric estimation of the transition probability matrix can be expressed as a function of the transition intensity in the form of a product integral[42]

$$\hat{\mathbf{P}}_{ij} = \prod\left(\mathbf{I} + d\hat{\boldsymbol{\Lambda}}\right) \tag{5}$$

where $\left[\Lambda_{ij}\right] = \hat{\mathbf{E}}$ is called the Aalen-Johansen estimator. The Aalen-Johansen estimator, which uses the mathematical framework of multivariate counting processes, reduces to simple empirical proportions for the complete data.

Let N_{hj}, i be a counting process that counts the number of i's patient direct transitions (without visiting another state in between) from state h to state j up to time t, where $h, j \in 1, \ldots, 4$ and $h \neq j$. Here, time t represents the time measured from the initial state, which is the complication-free state. Let Yj (t) refer to the number of individuals to be observed at risk in a state just prior to time t.

Then the conditional transition probability is specified using Equation (6).

$$\hat{p}_{hj}(t) = \frac{\Delta N_{hj}}{Y_j(t)} \tag{6}$$

where ΔN_{hj} gives the number of $h \rightarrow j$ transitions observed exactly at time t.

Kalbfleisch and Lawless[43] and later Kay R[44] described a general method for evaluating the likelihood of a general multistate model in continuous time, applicable to any form of transition matrix. The maximum likelihood that can be used to estimate the model parameters specified in Equation (3) is presented in equation (7) below. The partial likelihood function takes the product overall individuals and time points, as shown in equation (7).

$$L(\boldsymbol{\beta}) = \prod_{i \xrightarrow{\text{transition}} j} \prod_{k=1}^{n} \frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_{ij,k})}{\sum_{l \in R_i(t_{ij,k})} (\boldsymbol{\beta}' \mathbf{Z}_{ij,l})} \tag{7}$$

where $t_{hj,k}$ is the failure or censoring time of individual k for transition h→j, $d_{hj,k} = 1$, the

individual k has an event for transition $i \rightarrow j$, 0 otherwise, and *Rh(t)* is the risk set of state h at time t, that is, the set of individuals who are in state h at time *t* (t in this case is the time since the entry in state h). The estimate of the cumulative baseline hazard of transition i→j is the Nelson-Aalen estimate[45]

$$\Lambda_{hj,0(t)} = \sum_{k=1}^{n} \frac{d_{hj,k}}{\sum_{t_{hj,k} \le t} \sum_{l \in R_{h(t_{hj,k})}} \exp(\boldsymbol{\beta}' \mathbf{Z}_{ij,l})} \quad (8)$$

which is the proportion of the number of observed $h \rightarrow j$ transitions at tk to the number of individuals at risk in state h just prior to t_k with h, $j = r$, h $f = j$. In (4), the summation is over all event times tk, which are less than or equal to time t. Left truncation and right censoring are naturally accounted for in the denominator. The risk set in Equation (8) includes all individuals who have entered state h before time t and who have not yet moved out of state h again, provided that they have already entered the study (left-truncation) and are still under observation (right-censoring).

Results

Baseline Demographic and Clinical Characteristics of the Study Participants

The characteristics of 524 T2DM patients who were followed retrospectively from January 1, 2005, to December 31, 2017, are presented in Table 1 to assess their time for vascular complications or the date of death, whichever occurred first. All patients were free of T2DM-related complications when they first entered into follow-up. The median follow-up time was 7.4 with an interquartile range (IQR) of 6.01-10.02, implying that 50% of the patients had been under follow-up for at least 7.4 years. According to the results in Table1, the majority (66.27%) of patients with T2DM were men. Microvascular complications were present in 19% (n= 33) and 18% (n=66) of female and male diabetic patients, respectively; while 23% (n=40) of female and 21% (n=70) of male patients developed MaVasC's; 13% (n=23) and 18% (n= 62) of female and male diabetic patients, respectively died due to diabetes-related complications. In addition, 10% (n=18) of female and 11% (n=37) of male patients developed both MiVasC and MaVasC. The percentage of females with T2DM who experienced MaVasC was higher than that of males. The percentage of female patients with diabetes who experienced MaVasC was higher than that of their male counterparts. This revealed that female patients with diabetes were more likely than their male counterparts to develop chronic diabetic complications than their male counterparts. However, men were more likely to die from diabetes-related complications than their female counterparts. The majority (66.6%) of the patients were from urban areas. As for the distribution of diabetic patients regarding their vascular complication status and their death status, the percentage of rural dwellers who developed MiVasC, MaVasC, and died was consistently higher than that of urban dwellers. However, the effects of sex and residence on the differences in proportions were not statistically significant. More than half of the patients, 278(58.3%), had baseline blood pressure values of ≥140/90mmHg, whereas the remaining 199 (41.7%) patients had <140/90mmHg baseline blood pressure values. Moreover,

Table 1. Status of diabetic patients regarding vascular complication status and death

Covariates	All patients, n (%)	Status at the end		
		MiVasC, n (%)	MaVasC, n (%)	Died, n (%)
Sex				
Female	172 (33.73%)	33 (19%)	40 (23%)	23 (13%)
Male	338 (66.27%)	61 (18%)	7 0(21%)	62 (18%)
Residence				
Rural	170 (33.4%)	37 (22%)	36 (21%)	34 (20%)
Urban	279 (53.24%)	59 (17%)	69 (20%)	52 (15%)
BP category				
< 140/90 mmHg	199 (41.7%)	32 (11%)	30 (11%)	27 (10%)
≥ 140/90mmHg	278 (58.3%)	66 (33%)	73 (37%)	59 (30%)
Family history of diabetes				
Yes	353 (68.3%)	47 (13%)	56 (16%)	49 (14%)
No	164 (31.72%)	49 (30%)	52 (32%)	37 (23%)
serum creatine level				
> 1.2 mg/dL	64 (18.7%)	18 (28.13%)	24 (37.50%)	23 (35.94%)
≤ 1.2 mg/dL	289 (81.3%)	56 (20.07%)	86 (18.70%)	64 (13.94%)

Table 2. Characteristics of type II diabetic patients classified by patient state

Covariates	All Patients, Median(IQR)	States at the end, Median(IQR)		
		MiVasC	MaVasC	Dead
Age at diagnosis, years	52 (19)	58 (15)	60 (14)	65 (13)
Glucose level ,mg/dL	260 (154.00)	274.5 (108)	230 (110.5)	257 (112)
Systolic BP, mmHg	120 (20)	142 (23)	138 (30)	141.5 (31)
Diastolic BP, mmHg	82 (19)	84.5 (14)	83 (14)	89.5 (12)
Serum creatinine, mg/dL	0.88 (0.39)	0.98 (0.35)	0.96 (.63)	0.96 (.61)
Duration with DM, years	7.9 (4.01)	8.24 (5.4)	8.31 (3.4)	8.76 (.67)
HbA1c (%),	12.1 (3.45)	14.8 (8.4)	16 (6.1)	15.4 (4.13)
HDL, mg/dL	51 (27)	44 (39.75)	43 (25)	41 (11)
LDL, mg/dL	93 (55.50)	88 (36)	111 (53)	93 (42.75)
Triglycerides, mg/dL	4.2 (1.28)	137 (104)	136 (81)	146 (97.5)
Serum sodium, mEq/L	141 (10.50)	402 (199.5)	402 (0)	402 (266)
Serum potassium, mEq/L	4.2 (1.28)	4.89 (0.52)	4.89 (0.77)	4.55 (0.69)
Serum chloride, mEq/L	105.5 (11.5)	96.5 (8.63)	96.5 (0)	96.95 (11.5)

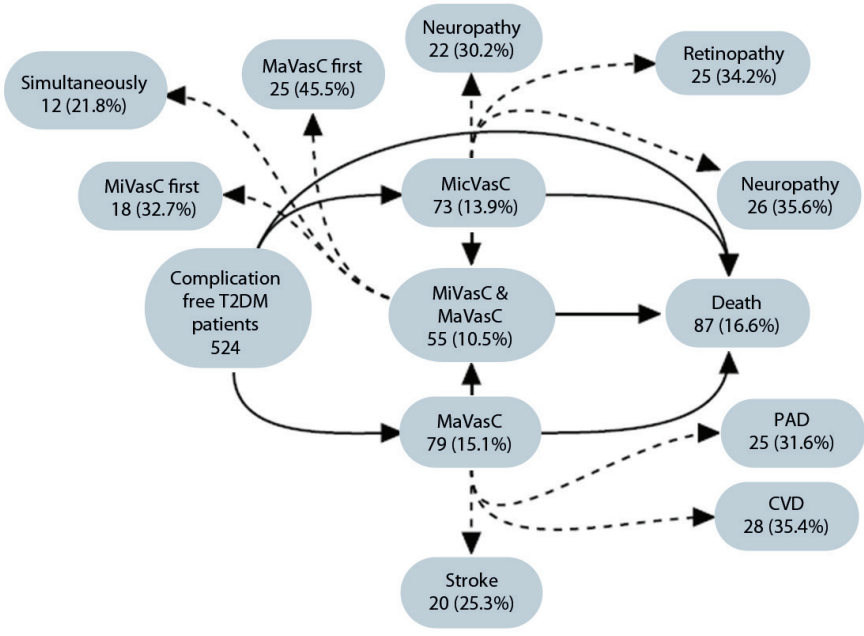

Figure 2. Possible transition and prevalence of diabetic complications among diabetic patients.
MiVasC, T2DM patients with micro-vascular complications; MaVasC, T2DM patients with macro-vascular complications; MiVasC & MaVasC, T2DM patients with both micro- and macro vascular complications; PAD, Peripheral artery disease; CHD, Chronic heart

the percentage of hypertensive T2DM patients (≥140/90mmHg) who developed MiVasC or MaVasC was higher than their non-hypertensive counterparts, suggesting that hypertensive diabetic patients were more likely to experience vascular complications than non-hypertensive patients. The log-rank test for all states (events) showed a statistically significant effect on blood pressure. From the 353 diabetes patients whose creatinine level was reported, 64(18.1%) patients were with elevated baseline serum creatinine (>1.2 mg/dL) and the rest 289 (81.9%) had normal serum creatinine (≤1.2 mg/dL).

Table 2 presents the median values of continuous predictors classified according to the clinical status of T2DM patients. Accordingly, the median age of the patients at the time of T2DM diagnosis was 52 years (IQR=19). The median triglyceride and creatinine levels were obtained to be 170 mg/dL (IQR=100.5 mg/dL) and 0.88 mg/dl (IQR = 0.39), respectively. The median fasting blood sugar and glycated hemoglobin percentage levels were 264mg/dl (IQR=154mg/dl) and 12.1(3.45), respectively.

Figure 2 provides information about the frequency of major micro- and macrovascular complications in patients with T2DM in the study area. According to the results in Figure 2, during an average follow-up period of 7.4 years, 152 (29%) of diabetic patients developed either MiVasC or MaVasC vascular complications, and 55(10.5%) of them diagnosed with both MiVasC and MaVasC. Specifically, during the follow-up time, T2DM patients had a probability of 0.14 and 0.15 to move into MiVasC and MaVasC states, respectively.

Moreover, from the 73 T2DM patients who have already developed vascular complications,

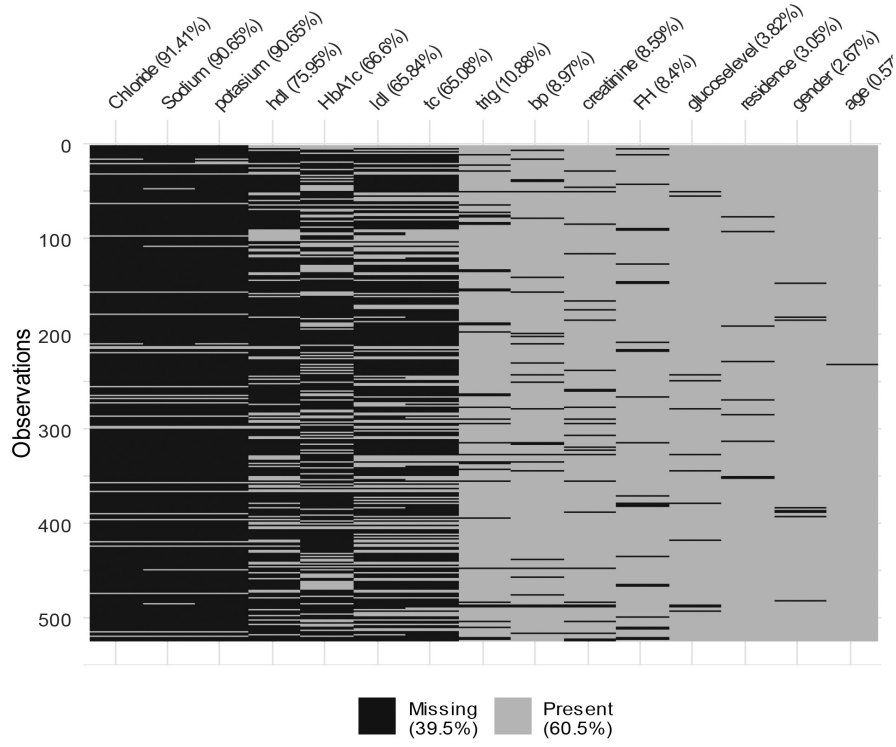

Figure 3. Heatmap visualization of missing data for the T2DM data

18 of them developed macrovascular complications, revealing that T2DM patients who have already developed MiVaC had a 0.25 probability to develop MaVasC following MiVasC. Similarly, from the 79 T2DM patients who have already developed MaVasC had, 25 of them experienced MiVasC, showing that T2DM patients had a 0.32 chance to experience MiVasC disease following MaVasC. Among T2DM patients who experienced microvascular complications, the proportion of retinopathy, nephropathy, and neuropathy was 34.2%,30.2%, and 35.6% respectively. On the other hand, the cases of stroke, chronic heart disease (CHD), and peripheral artery disease (PAD), respectively, accounted for 20.3%, 35.4%, and 31.6% of the total macrovascular cases. Furthermore, the probability of T2DM patients dying due to other causes (dying without developing vascular complications) is 0.15. Similarly, T2DM had 0.15 and 0.27 probabilities of dying once they experienced microvascular and macrovascular complications, respectively.

Exploring Patterns of Missingness

The heatmap produced using vis_miss() from Visdat package[46] provides a pattern and percentage summary of missingness for each variable or column is shown in Figure 3. This plot helps the visualization of the pattern of percentages of missing values distribution.[47]
It can be seen from Figure 3 that 8 covariates have a missing value with a very small proportion, whereas 7 variables including fasting serum lipids (HDL, LDL, total cholesterol), glycated hemoglobin percentage (HbA1%), and serum electrolyte (potassium, sodium, chlorine levels) have an extraordinary

proportion of missing data. The proportion of missing data should not be used to guide decisions on multiple imputation.[48] Therefore, we decided to take out the seven covariates and rather work with the other covariates that have less proportion missing data and can be successfully imputed.[29] Imputation can be more precise if other variables are non-missing for those cases that are to be imputed.[49]

Nonparametric Analysis of Transition Intensity Rates and Probability

Figure 4 presents the estimated transition probabilities from all starting states to all possible states, between the starting time t = 0 and all event times successively. The first panel (a) provides a plot of transition probabilities from state 1. It is observed from this figure that the estimated probability of staying in the same state decreases with increasing time. A sharp drop in the probability of being remaining in the same state (state) reveals that the probability of staying free from diabetes-related complications during their follow-up period decreases with time. The probability of remaining free from any complication is estimated at 0.76 at 5 years and 0.4 at 10 years. Moreover, the median complication-free time was 8.85 years, where 50% of T2DM patients are estimated to survive being free from any diabetes-related complications for a duration of 8.85 years since their first time diagnosed with diabetes. The probability of the T2DM patient moving into state 4 (death state) increases with time. More specifically, the probability of T2DM patients dying due to diabetes-related complications was estimated to be 0.02 and 0.039 at 5 years and 10 years, respectively. It was also revealed in the figure that the patient had an increased probability to move into a micro-and macrovascular complication

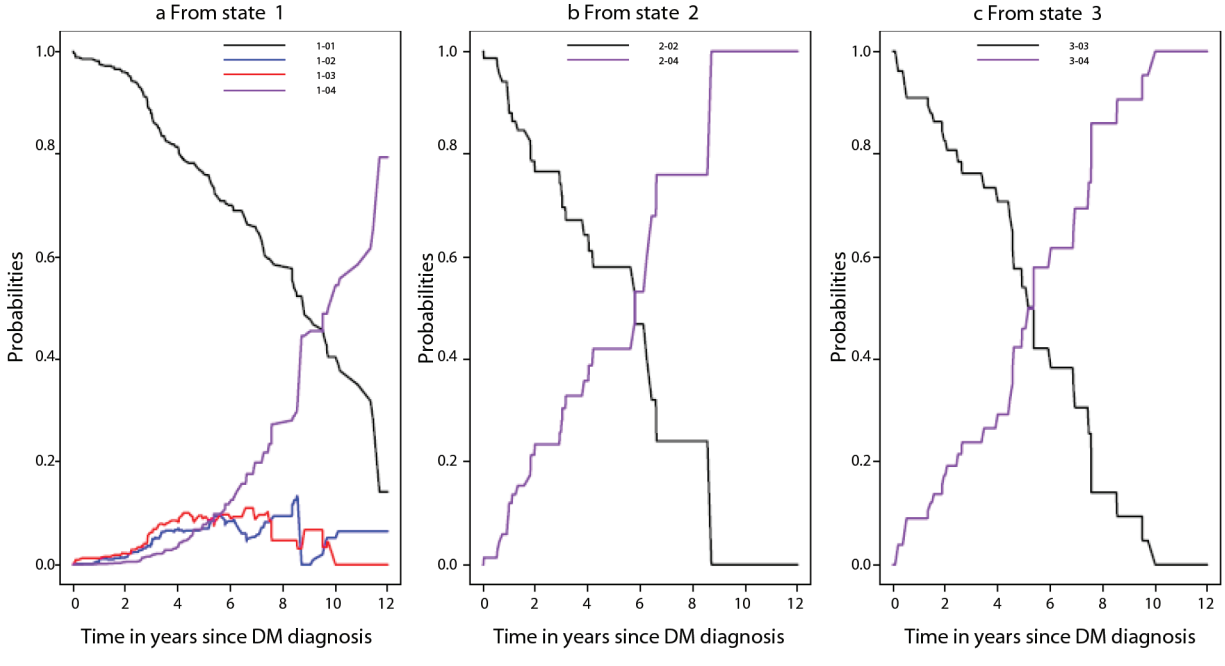

Figure 4. Prediction of transition probability with time to vascular complications for diabetic patients starting from each state at t = 0, where 1=complication free; 2= microvascular complications; 3= macrovascular complications; 4= death

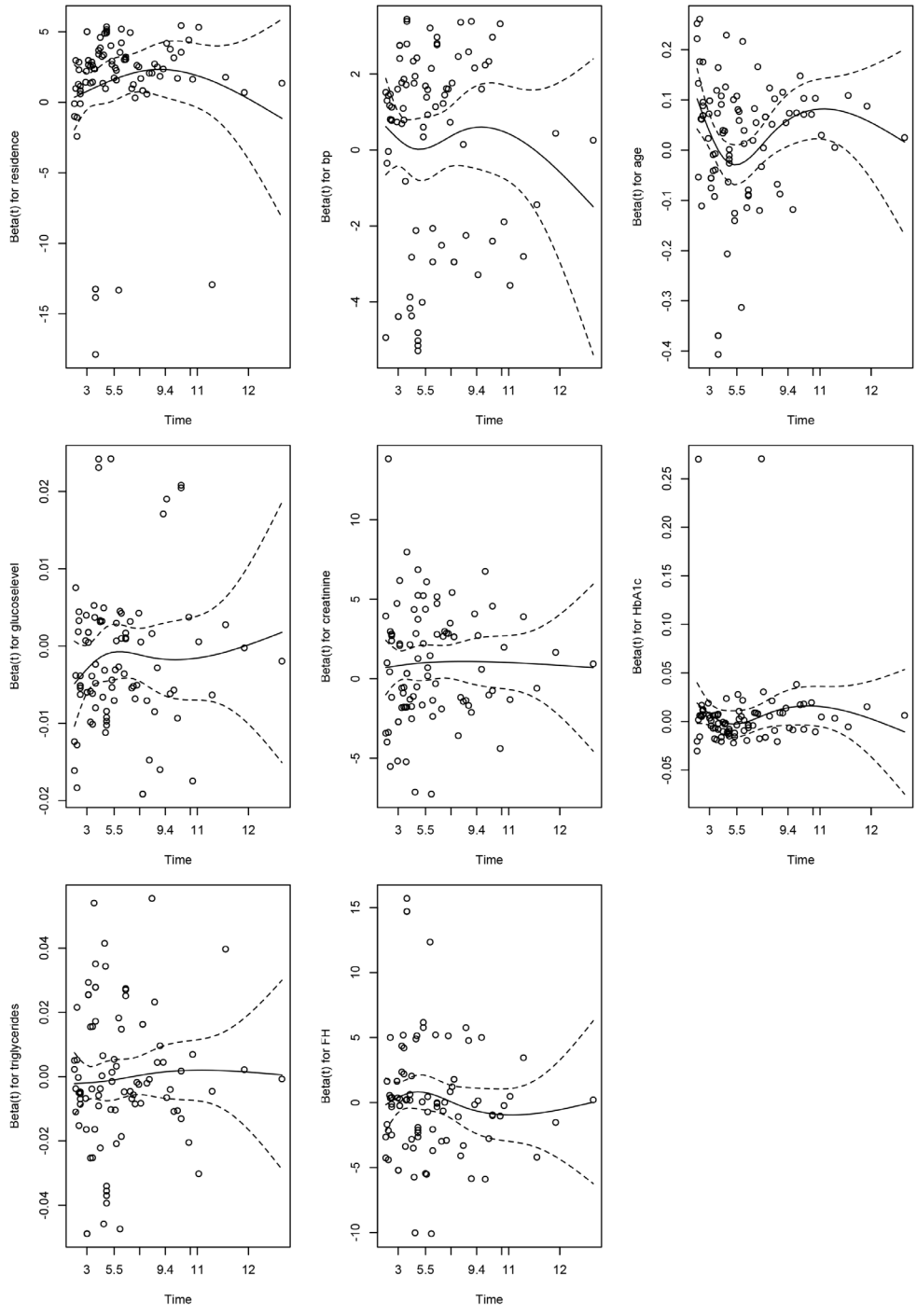

Figure 5. Plot of a scaled Schoenfeld residuals for testing the proportionality assumption

state from the beginning of the follow-up. However, the curve becomes stable for the fact that patients will understand the danger of the disease as time passes and modify their lifestyle, and they improve their adherence to the drug. The second and the third panels of the figure (b & c) provide the transition probabilities from state 1 and state 2, respectively. Both figures

Table 3. Table showing the proportional hazard assumption check

Covariates	Chisq	df	P	Does proportionality assumption hold?
Residence (Urban)	7.39	1	0.01	No
Blood pressure(<140/90)	1.01	1	0.32	Yes
Age at diagnosis	0.95	1	0.33	Yes
Baseline glucose level	1.71	1	0.19	Yes
Creatinine level	3.85	1	0.05	No
HbA1c (%)	1.15	1	0.28	Yes
Triglycerides	0.55	1	0.46	Yes
Family history of diabetics	1.13	1	0.29	Yes
GLOBAL	16.73	8	0.03	No

revealed that the probability of staying in the same state monotonically decreased with time, whereas the probability that DM patients move to the next state increased with time.

Multistate Model Results

Model Building and Assumption Assessment
After imputation, a multistate Markov model was fitted to the imputed data. The four states of the multistate Markov model include complication-free, MiVasC, MaVasC, and death (Figure 1). The death state is an absorbing state. To begin the model fitting process, we used the likelihood ratio test if the fitness of the model was improved after incorporating all sojourn times in different states. The likelihood ratio test showed that the fitness of the model to explain the data was significantly improved after incorporating the sojourn times, revealing that the Markov assumption was violated, and hence the transition hazards/intensities dependent on the time spent in the current state. To improve the prediction of the transition probabilities, baseline predictors were considered. To select important predictors, covariates that were statistically significant at 10% level of significance in the univariable model with at least one transition were included in the higher (multivariable). Accordingly, age at the onset of diabetes, residence (rural or urban); blood pressure category (<140/90mmHg or ≥ 140/90 mmHg), family history of diabetes, triglyceride level, baseline hemoglobin level, and serum creatinine level were significant at the fixed level of significance in the univalent model. Hence, the remaining covariates were not included in the final model. The assumption of proportional hazard or time homogeneity for the transitions needs to be checked to see if it is reasonably fulfilled for our data. To check the model assumption, both formal tests and graphical plots can be used. For this analysis, we used graphical plots (as shown in Figure 5) to check the assumption. The plots look fairly flat, revealing that the assumption of proportionality is not violated provided that the plots are decidedly different from flat.
However, a formal test using the cox.zph

Table 4. Hazard Ratios and P-values for the final stratified Cox model

Covariates	Possible transitions									
	1− > 2		1− > 3		1− > 4		2− > 4		3− > 4	
	HR	P	HR	P	HR	P	HR	P	HR	P
Residence Rural (ref)										
Urban	0.31	0.85	4	0.68	2.1	0.01	3.44	0.47	1.31	0.05
Triglycerides	0.98	0.46	1	0.12	1	0.7	1.97	<0.001	1	0.64
Age	1	0.93	1.06	0.35	1.09	0.03	2.1	0.01	3.14	0.01
Blood pressure category										
Hypertensive	6.24	0.04	5.2	0.008	2.13	0.019	10.81	0.04	0.3	0.37
Family history of diabetic										
Yes	1.94	0.6	1.95	0.08	1.26	0.09	1.81	0.4	4.3	0.03
Creatinine	1.4	0.66	2.32	0.18	1.07	0.95	214.469	0.044	1.86	0.67
Baseline glucose level	1.007	0.002	0.996	0.106	0.992	0.155	1.004	0.343	1.003	0.937
Duration with T2DM	-	-	-	-	-	-	1.7 3	<0.001	14.01	0.08

1: Complication free state.
2: Microvascular complication state.
3: Macrovascular complication state.
4: Death.

function in the survival package R software was used to generate the p values of the tests based on the scaled Schoenfeld residuals for non-proportional hazard assessment and the output of the test presented in Table 3. Based on the output generated, all tests except for two covariates (residence and baseline creatine level) have shown non-violation of the proportional hazards assumption. The fitness, as well as the violation proportionality assumption, were improved when a model whose transition intensities to death (transitions specified by 4) are assumed to be proportional to the baseline hazards.

Checking the Markove Assumption

The duration of patients with T2DM (sojourn time in the diabetic complication-free state) was included as a prognostic covariate to test if it has a statistical significant relationship with the transition to death state from MiVasC or MaVas C. The fact that the sojourn time of patients had a significant association with the transition of the next transition of the model revealed an evidence that the Markov assumption is not valid. This calls for semi-Markov models as an alternative model for the analysis. The model's output is presented in Table 4 in which P-values and hazard ratio (HR) values of the predictors having statistically significant association (at 5% significance level) with the states are italicized. According to the model's results, keeping the effects of predictors in the model constant, an increased age at the time of diabetic diagnosis was associated with increased transition intensity into the death state. More specifically, a year increase in the age of patients at T2DM diagnosis corresponds to a 9% increase in the death of T2DM patients without experiencing any

complications (or complications might not be detected). Moreover, a year increase in the age of T2DM patients at the time of diagnosis increases the risk of dying by a factor of 2.1 and 3.14 times for patients with microvascular and macrovascular, respectively, showing that an increased age at diagnosis corresponded with the risk of earlier vascular complications onset and death. The effect on the event of death was pronounced more after T2DM patients experienced vascular complications of either type. However, age at diagnosis showed no significant difference in the transition rate between type II diabetic patients in micro-or macrovascular complication states. Similarly, baseline blood glucose level was a statically significant predictor of the transition of patients into a microvascular complication state such that an increased baseline glucose level had an increased risk of developing microvascular complications. Furthermore, rural dwellers were more likely to move into the death state from the complication-free state (HR=2.1, P=0.01) and from the microvascular complication state (HR=1.31, P=0.05). Type II diabetic patients with a family history of diabetes had more likely to move from the macrovascular complication state into the death state. The presence of hypertension is a risk factor for the development of vascular complications or death. Hypertensive patients, for example, were 2.13 times more likely to move to the death state from the complication-free state compared to non-hypertensive DM patients after sole adjustment of predictors in the model. Besides, from the final model, we summarize that the time spent in the complication-free state (duration of diabetes) has a prognostic impact on the transition from macrovascular complication to death.

Discussion

Multistate models provide a flexible framework for understanding clinical events under consideration of the disease process as a whole, not only focusing on one single endpoint, like the classical Cox regression model.[23] In the applications of multistate, the dependence on the history of the process is negligible and gives sense to assume the Markov property, where the state sojourn times follow the exponential distribution with a constant hazard rate.[50] However, in the study of chronic disease, Markov assumption is unrealistic. For instance, in the study of T2DM disease progression, the disease stages may not follow a constant hazard rate. As a result, Semi-Markov models are applied for modeling a variety of phenomena in different areas,[51] such as the progression of breast cancer,[29] engineering,[52] and different chronic disease progression.[53]

In this study, we used Semi-Markov model to model multistate chronic disease transitions by taking into consideration of the length of time that patients have stayed in a certain disease state, which is clinically more plausible. Semi-Markov models are applicable to degenerative diseases, while Markov models are applicable to epidemic diseases.[54]

Therefore, a semi-Markov multistate model was adapted using the clock-reset approach to study the effects of different covariates on transit to the next state. During an average of 7.4 years follow-up period, nearly 28% of diabetic patients developed vascular complications of either type, which

slightly higher than what has been previously reported by Wolde et al.[13] based on data from the University of Gondar Referral Hospital, Ethiopia, but lower than the study conducted in China.[55] The difference in the length of the follow-up period and the age groups of the study participants could be the reason for such discrepancy. Duration of diabetic illness was associated with an increased risk of death. The same findings were reported by.[56, 57] An increased age at diagnosis corresponded with the increased likelihood of death. This feeding is in line with the findings by Kidanie, Alem,[12] Wolde, Atsedeweyen,[13] Zoungas, Woodward,[55] which was conducted in Ethiopia and Li, Chattopadhyay[56] in China. This could be due to the fact that aging results in relative insulin resistance, cell dysfunction, altered glucose metabolism, associated hormonal changes, and decreased physical activity,[58, 59] which in turn leads to death.[56] In our study, women were more likely than men to develop chronic diabetic complications, although sex was not a significant predictor of diabetic complication status. This is because women are subjected to a partial loss of estrogen protection normally present in the premenopausal state.[60] In this study, being hypertensive was higher in the transition into microvascular and macrovascular complications as well as death state. This result is supported by Li, Chattopadhyay,[56] Agrawal, Ola,[61] and Tracey, McHugh.[62]

Conclusion

Multistate models are good statistical methods for exploring predictors and multiple interrelated events by controlling their dependence. In this study, the semi-Markov model fitted the data well and could be taken as a convenient model for the analysis of time to diabetes-related complications or death in this particular data. Moreover, blood pressure, age at diagnosis of diabetes, and diabetes duration are were predictors of vascular complications and death. The percentage of diabetes-related complications and their consequences, clearly death, is not small, although the study needs to be confirmed by using an adequate sample. Thus, appropriate prevention and control strategies should be in place to limit further progress and impact. The current results could also be taken as a baseline for others who aim to explore the occurrence of vascular complications as a function of different risk factors in individuals with T2DM.

Conflict of interest

The authors declare that none of us has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

Ethics Approval and Consent to Participate

This study was approved by the Debre Markos University Research Ethics Committee and the medical director of the Hospital. As the study was retrospective, informed consent was not obtained from the study participants, but the anonymity and the confidentiality of the data were assured.

Data availability statement

The data set analyzed in this study is available

from the corresponding author on reasonable request.

Funding

This work was financially supported by Debre Markos University.

Acknowledgments

The authors are thankful to Debere Markos Referral Hospital for the permission to use the data and staff members for their support in extracting the information from the patient's medical cards. We are also grateful to Debre Markos University for financial support.

Authors' Contributions

MGA contributed to the acquisition of data, study concept, and design of the statistical methodology, and performed the analysis and interpretation of the data, as well as wrote the first draft of the manuscript and ASI contributed to the study on critical revision of the manuscript. The authors read and approved the final manuscript.

References

1. Organization, W.H, Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. 1999, World Health Organization.
2. Association, A.D, Diagnosis and classification of diabetes mellitus. Diabetes care, 2014. 37(Supplement 1): p. S81-S90.
3. Baena-Díez, J.M., et al., Risk of cause-specific death in individuals with diabetes: a competing risks analysis. Diabetes care, 2016. 39(11): p. 1987-1995.
4. Shaw, J.E., R.A. Sicree, and P.Z. Zimmet, Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes research and clinical practice, 2010. 87(1): p. 4-14.
5. Cho, N., et al., IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes research and clinical practice, 2018. 138: p. 271-281.
6. Ogurtsova, K., et al., IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice, 2017. 128: p. 40-50.
7. Organization, W.H., Global report on diabetes: executive summary. 2016, World Health Organization.
8. Skyler, J.S., et al., Differentiation of diabetes by pathophysiology, natural history, and prognosis. Diabetes, 2017. 66(2): p. 241-255.
9. Alemu, F., Prevalence of diabetes mellitus disease and its association with level of education among adult patients attending at Dilla Referral Hospital, Ethiopia. J Diabetes Metab, 2015. 6(4): p. 1-5.
10. Patil, P.D., et al., Past and current perspective on new therapeutic targets for Type-II diabetes. Drug design, development and therapy, 2017. 11: p. 1567.
11. Konstantinos, P., et al., Complications of diabetes 2016. Journal of diabetes research, 2016.
12. Kidanie, B.B., et al., Determinants of Diabetic Complication Among Adult Diabetic Patients in Debre Markos Referral Hospital, Northwest Ethiopia, 2018: Unmatched Case Control Study. Diabetes, metabolic syndrome and obesity: targets and therapy, 2020. 13: p.

237.
13. Wolde, H.F., et al., Predictors of vascular complications among type 2 diabetes mellitus patients at University of Gondar Referral Hospital: a retrospective follow-up study. BMC endocrine disorders, 2018. 18(1): p. 1-8.
14. de Wreede, L.C., M. Fiocco, and H. Putter, mstate: an R package for the analysis of competing risks and multi-state models. Journal of statistical software, 2011. 38(7): p. 1-30.
15. Machado, R.J. and A. van den Hout, Flexible multistate models for interval-censored data: Specification, estimation, and an application to ageing research. Statistics in medicine, 2018. 37(10): p. 1636-1649.
16. Meira-Machado, L., et al., Multi-state models for the analysis of time-to-event data. Statistical methods in medical research, 2009. 18(2): p. 195-222.
17. Putter, H., Tutorial in biostatistics: Competing risks and multi-state models Analyses using the mstate package. Companion file for the mstate package, 2011.
18. Therneau, T., C. Crowson, and E. Atkinson, Multi-state models and competing risks. CRAN-R (https://cran. r-project. org/web/packages/survival/vignettes/compete. pdf), 2020.
19. Foucher, Y., et al., A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. Statistical methods in medical research, 2010. 19(2): p. 127-145.
20. Kalbfleisch, J.D. and R.L. Prentice, The statistical analysis of failure time data. Vol. 360. 2011: John Wiley & Sons.
21. Aalen, O.O. and S. Johansen, An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics, 1978: p. 141-150.
22. Abner, E.L., R.J. Charnigo, and R.J. Kryscio, Markov chains and semi-Markov models in time-to-event analysis. Journal of biometrics & biostatistics, 2013(e001): p. 19522.
23. Eulenburg, C., et al., A systematic model specification procedure for an illness-death model without recovery. PloS one, 2015. 10(4): p. e0123489.
24. Le-Rademacher, J.G., et al., Application of multi-state models in cancer clinical trials. Clinical Trials, 2018. 15(5): p. 489-498.
25. Developer Core Team, R., R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2019.
26. Therneau, T. and T. Lumley, R survival package. 2013.
27. Jerez, J.M., et al., Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artificial intelligence in medicine, 2010. 50(2): p. 105-115.
28. Nakagawa, S., Missing data: mechanisms, methods and messages. Ecological statistics: Contemporary theory and application, 2015: p. 81-105.
29. Kotzé, L., Markov modelling of disease progression in the presence of missing covariates. 2019, Stellenbosch: Stellenbosch University.
30. Honaker, J., G. King, and M. Blackwell, Amelia II: A program for missing data. Journal of statistical software, 2011. 45(7): p. 1-47.
31. Buuren, S.v. and K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R. Journal of statistical software, 2010: p. 1-68.

32. Kowarik, A. and M. Templ, Imputation with the R Package VIM. Journal of Statistical Software, 2016. 74(7): p. 1-16.
33. Stekhoven, D.J. and M.D.J. Stekhoven, Package 'missForest'. 2012.
34. Stekhoven, D.J. and P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics, 2012. 28(1): p. 112-118.
35. Tang, F., Random Forest Missing Data Approaches. 2017: University of Miami.
36. Bjerg, L., et al., Development of microvascular complications and effect of concurrent risk factors in type 1 diabetes: a multistate model from an observational clinical cohort study. Diabetes Care, 2018. 41(11): p. 2297-2305.
37. Anderson, D. and K. Burnham, Model selection and multi-model inference. Second. NY: Springer-Verlag, 2004. 63(2020): p. 10.
38. Li, X. and M. Fiocco, Estimation for Non-Markov Multi-states Models. 2014.
39. Andersen, P.K., L.S. Hansen, and N. Keiding, Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. Scandinavian Journal of Statistics, 1991: p. 153-167.
40. Zare, A., et al., Assessing Markov and time homogeneity assumptions in multi-state models: application in patients with gastric cancer undergoing surgery in the Iran cancer institute. Asian Pacific Journal of Cancer Prevention, 2014. 15(1): p. 441-447.
41. Conlon, A., J. Taylor, and D. Sargent, Multi-state models for colon cancer recurrence and death with a cured fraction. Statistics in medicine, 2014. 33(10): p. 1750-1766.
42. Andersen, P.K., et al., Statistical models based on counting processes. 2012: Springer Science & Business Media.
43. Kalbfleisch, J. and J.F. Lawless, The analysis of panel data under a Markov assumption. Journal of the american statistical association, 1985. 80(392): p. 863-871.
44. Kay, R., A Markov model for analysing cancer markers and disease states in survival studies. Biometrics, 1986: p. 855-865.
45. Aalen, O., Nonparametric inference for a family of counting processes. The Annals of Statistics, 1978: p. 701-726.
46. Tierney, N., visdat: Visualising whole data frames. Journal of Open Source Software, 2017. 2(16): p. 355.
47. Ghazali, S.M., N. Shaadan, and Z. Idrus, Missing data exploration in air quality data set using R-package data visualisation tools. Bulletin of Electrical Engineering and Informatics, 2020. 9(2): p. 755-763.
48. Madley-Dowd, P., et al., The proportion of missing data should not be used to guide decisions on multiple imputation. Journal of clinical epidemiology, 2019. 110: p. 63-73.
49. Jakobsen, J.C., et al., When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. BMC medical research methodology, 2017. 17(1): p. 1-10.
50. Asanjarani, A., B. Liquet, and Y. Nazarathy, Estimation of semi-Markov multi-state models: a comparison of the sojourn times and transition intensities approaches. The International Journal of Biostatistics, 2020. 1(ahead-of-print).
51. Janssen, J., Semi-Markov models: theory and applications. 2013: Springer Science & Business Media.
52. Świderski, A., et al., Evaluation of Machinery Readiness Using Semi-Markov

Processes. Applied Sciences, 2020. 10(4): p. 1541.
53. Liu, P., L. Liao, and J. Liu. Chronic Disease Progression Modeling using Semi-Markov Model with Noisy Observations. in IIE Annual Conference. Proceedings. 2015. Institute of Industrial and Systems Engineers (IISE).
54. Brookmeyer, R. and N. Abdalla, Multistate models and lifetime risk estimation: Application to Alzheimer's disease. Statistics in medicine, 2019. 38(9): p. 1558-1565.
55. Zoungas, S., et al., Impact of age, age at diagnosis and duration of diabetes on the risk of macrovascular and microvascular complications and death in type 2 diabetes. Diabetologia, 2014. 57(12): p. 2465-2474.
56. Li, J., et al., Prevalence and associated factors of vascular complications among inpatients with type 2 diabetes: A retrospective database study at a tertiary care department, Ningbo, China. PloS one, 2020. 15(6): p. e0235161.
57. Rangel, E.B., C.O. Rodrigues, and J.R. De Sa, Micro-and macrovascular complications in diabetes mellitus: preclinical and clinical studies. 2019, Hindawi.
58. Chia, C.W., J.M. Egan, and L. Ferrucci, Age-related changes in glucose metabolism, hyperglycemia, and cardiovascular risk. Circulation research, 2018. 123(7): p. 886-904.
59. Kalyani, R.R. and J.M. Egan, Diabetes and altered glucose metabolism with aging. Endocrinology and Metabolism Clinics, 2013. 42(2): p. 333-347.
60. Awa, W.L., et al., Type 2 diabetes from pediatric to geriatric age: analysis of gender and obesity among 120183 patients from the German/Austrian DPV database. European Journal of Endocrinology, 2012. 167(2): p. 245.
61. Agrawal, R., et al., Prevalence of micro and macrovascular complications and their risk factors in type-2 diabetes mellitus. JAPI, 2014. 62: p. 505.
62. Tracey, M.L., et al., Risk factors for macro-and microvascular complications among older adults with diagnosed type 2 diabetes: findings from the Irish longitudinal study on ageing. Journal of diabetes research, 2016. 2016.