



Published in final edited form as:

Inf Fusion. 2012 April 1; 13(2): 137–145. doi:10.1016/j.inffus.2010.08.003.

Use of Secondary Data to Estimate Instantaneous Model Parameters of Diabetic Heart Disease: Lemonade Method

Wen Ye, Deanna JM Isaman, and Jacob Barhak

Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029

Abstract

With the increasing burden of chronic diseases on the health care system, Markov-type models are becoming popular to predict the long-term outcomes of early intervention and to guide disease management. However, statisticians have not been actively involved in the development of these models. Typically, the models are developed by using secondary data analysis to find a single “best” study to estimate each transition in the model. However, due to the nature of secondary data analysis, there frequently are discrepancies between the theoretical model and the design of the studies being used. This paper illustrates a likelihood approach to correctly model the design of clinical studies under the conditions where 1) the theoretical model may include an instantaneous state of distinct interest to the researchers, and 2) the study design may be such that study data can not be used to estimate a single parameter in the theoretical model of interest. For example, a study may ignore intermediary stages of disease. Using our approach, not only can we accommodate the two conditions above, but more than one study may be used to estimate model parameters. In the spirit of “If life gives you lemon, make lemonade”, we call this method “Lemonade Method”. Simulation studies are carried out to evaluate the finite sample property of this method. In addition, the method is demonstrated through application to a model of heart disease in diabetes.

Keywords

diabetes; disease modeling; chronic disease; designed absorption; multi-state model; meta-analysis

1. Introduction

With the escalating costs of health care, health care managers and policy makers are increasingly interested in preventive approaches to disease management. This has lead many researchers to develop Markov-type models of the natural history of chronic diseases in order to predict the long-term benefits of intervention [1].

The simplest Markov chain (discrete-state discrete-time Markov chain) is that in which there are a finite number of states and a finite number of equidistant time points at which observations are made, the chain is first-order, and the transition probability is the same for

© 2010 Elsevier B.V. All rights reserved.

Phone: +1 (734) 647-5686, Fax: +1 (734) 763-2215, University of Michigan, Ann Arbor, wye@umich.edu , djmisaman@umich.edu , jbarhak@umich.edu .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

each time interval. Such a chain is described by the initial state and the set of transition probabilities; namely, the conditional probability of going into each state, given the immediate preceding state. Modeling a disease process with such a model has the advantage of being intuitive enough to allow a relatively simple model formulation, yet allow describing complex processes.

For example, we are interested in a model of heart disease in people with diabetes. This model is one part of a discrete-state discrete-time Markov simulation model of the natural history of diabetes described previously [2]. The theoretical model is illustrated below in Figure 1, where boxes indicate states in the model and arrows indicate possible transitions between the states. Our theoretical model for heart disease has 5 states, ordered 0 to 4 respectively for no cardiovascular disease (CVD), angina, myocardial infarction (MI), history of MI (Hx MI), and death. A myocardial infarction is defined as an instantaneous state such that patients progress through MI instantaneously and either die or survive (and enter a state called “history of MI”). The parameters of interest are denoted by the initial and terminal states of transfer, for example q_{01} denotes the probability of progression from state 0 to state 1.

In order to estimate transition probabilities in such a model, a longitudinal study of the natural history of diabetes would provide ideal information. However, due to the long duration of diabetes, very few longitudinal studies exist that measure multiple states in the theoretical model. In most applications, there is no single study from which all parameters are estimable. Therefore, these chronic disease models are constructed from secondary data analysis of the clinical literature [3]. Summary statistics reported by a variety of studies in the clinical literature are extracted and used as point estimates for transitions in the theoretical model, often one for each transition. Alternatively, model calibrations are used to select parameter values by perturbing model parameters one at a time to reproduce expected or known results in calibration data [4]. Another alternative estimation method for a non state transition model focuses on estimation from person specific data [5]. Such an approach, however, requires accessibility to such data, which is often not publicly available and therefore constitutes a difficulty.

There were two challenges faced when developing this model: 1) the existence of an instantaneous state (e.g., MI), and 2) the use of secondary data which do not match the design of our theoretical model.

Since sojourn time for most transitions is much longer than one year, and since information from studies is usually reported at the time scale of years, diabetes disease models [2][6][7] often use one year as the time interval for discrete transition probabilities. However, some transitions may occur on a faster time scale. In the diabetic heart disease model a patient may die immediately or within a few months after experiencing myocardial infarction (MI). However, when MI is modeled as a non-instantaneous state, a subject who experiences MI during the current year can not die till next year. In addition, we wish to model the state MI as instantaneous for the sake of an economic analysis that weights the occurrence of an MI separately from the annual costs of surviving an MI, and dying from MI. Therefore it is necessary for us to estimate the transition parameter for angina to MI separately from MI to surviving MI and MI to death. Because our underlying model is structured as a discrete-time Markov chain, we had to modify the transition matrix to accommodate this exception to the discrete-time framework. In addition, our approach to the problem of instantaneous states had to be general enough to accommodate the use of secondary data.

The use of secondary data presented difficulties in our model because the studies providing data are not always designed consistently with respect to our theoretical model. This failure

of data to perfectly match the design of the model is a very common problem in secondary data analysis. For example, the United Kingdom Prospective Diabetes Study (UKPDS), a prospective, population-based longitudinal study, reported progression from healthy diabetic patients with no history of angina or MI (state 0 in our theoretical model) to the first coronary heart disease (CHD) event [8]. CHD is defined in [8] as “the occurrence of fatal or non-fatal MI or sudden death”. We assume that the CHD event as defined in UKPDS matches our definition of MI (state 2 in our theoretical model). This transition studied by the UKPDS is not explicit in our theoretical model, which provides two paths to the first MI event as shown in Figure 1: one that passes through angina and one that does not. As such, the single point estimate provided by the UKPDS is a function of four transition parameters ($q_{01}, q_{02}, q_{12}, q_{14}$) under our theoretical model. Thus, the UKPDS cannot provide us with an estimate for any one of these parameters of interest. Although the four parameters are not estimable using the UKPDS alone, we wish to use the UKPDS along with other studies to estimate our disease model. Isaman and colleagues call this valuable, but confounding data as *augmentary data* [3], and presented a likelihood based method to estimate all parameters in the model simultaneously using secondary data which may or may not match the design of the theoretical model for the complete disease process. We name this estimation approach the Lemonade Method in the spirit of “when study data give you lemons, make lemonade”. To our knowledge, to date, there is no other method capable of estimating parameters in a state transition model while handling instantaneous states, and using augmentary data that can be found in the literature.

The following sections will extend the Lemonade Method in [3] to include both instantaneous states and indirect data into our model of heart disease; will use simulation to investigate the finite-sample behavior of our approach using a simple example; and apply this new method for estimating parameters in a model of heart disease in diabetes (Figure 1).

2. The Methods

We first introduce some matrix manipulation to construct the transition probability matrix representing the theoretical model that takes into account the instantaneous transitions between states. Then we will describe the likelihood method for handling secondary data when an instantaneous state is present in the theoretical model.

2.1. Handling Instantaneous State

To accommodate instantaneous states in the framework of a discrete-state discrete-time Markov model, we introduce the following notation. Let

N denote the number of nodes (states) in the theoretical model,

q_{ij} denote the transition probability from state i to state j (transition probability in unit time if state i is non-instantaneous state, and instantaneous transition probability if state i is instantaneous state),

Ψ denote the set of the instantaneous states in the model,

\mathbf{E} denote an $N \times N$ matrix of instantaneous-state transitions such that

$$e_{ij} = \begin{cases} 1 & i \notin \Psi, \quad i=j \\ q_{ij} & i \in \Psi, \quad i \neq j \\ 0 & \text{otherwise} \end{cases}$$

(e.g., for the model in Figure 1 this matrix is,

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ q_{20} & q_{21} & 0 & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{23} & 1 - q_{23} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where q_{20} and q_{21} are both 0, and $q_{24}=1-q_{23}$ according to the structure of the model),

\mathbf{R} denote an $N \times N$ matrix of non-instantaneous state transitions such that

$$r_{ij} = \begin{cases} 0 & i \in \Psi, \quad i \neq j \\ q_{ij} & i \notin \Psi, \quad i \neq j \\ 1 - \sum_{b \neq i} r_{ib} & i = j \end{cases}$$

(e.g., for the model in Figure 1 this matrix is

$$\mathbf{R} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 - q_{12} - q_{14} & q_{12} & 0 & q_{14} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

), and

\mathbf{P} denote the $N \times N$ transition matrix representing the theoretical model that takes into account the instantaneous transitions between states, with elements $\{P\}_{ij} = \pi_{ij}$. As we will show below, \mathbf{P} can be derived with \mathbf{E} and \mathbf{R} matrices. Therefore, π_{ij} 's are functions of q_{ij} 's.

To understand why \mathbf{E} and \mathbf{R} matrices are constructed as described above, we can think of the transition in one time unit as if there are two steps. The \mathbf{R} matrix holds the transition probability in the first step and \mathbf{E} matrix holds the transitions probabilities in the second step. In the first step, all the transitions from non-instantaneous state take place, and those subjects who have transited to the instantaneous state stay in the instantaneous state temporarily. In this step, no transition emanates from the instantaneous state. That is why in the \mathbf{R} matrix, all elements on the rows corresponding to non-instantaneous states take the values of the corresponding q_{ij} 's. On the row for the instantaneous state, the diagonal element is 1 and the rest of the elements are zero (no transition emanating from the instantaneous state). In the second step, all subjects in non-instantaneous states remain in the same state, and all subjects that temporarily stay in the instantaneous state after the first step transit to other states. Therefore, in the \mathbf{E} matrix, for all rows corresponding to non-instantaneous state, by definition the diagonal elements are 1 and the rest of the row elements are zero, no transition is allowed to emanate from these states. For a row corresponding to an instantaneous state, elements take the corresponding q_{ij} 's. Note that by the definition of an instantaneous state, a subject may not remain in an instantaneous state, and therefore the diagonal element in the instantaneous state matrix for an instantaneous state is always zero.

Assuming no loops among instantaneous states or consecutive instantaneous states, we consider the three following type of model structures in terms of instantaneous states.

First, when there are no consecutive instantaneous states, then, by construction, the product of these two matrices \mathbf{RE} incorporates the transitions from both instantaneous and non-instantaneous states.

Second, if two or more instantaneous states exist consecutively in the system, then two or more instantaneous transitions may occur in one time unit. In this situation, $\mathbf{P}=\mathbf{RE}^c$ where c is the number of instantaneous states in the model.

Finally, more generally, if more than one queue of consecutive instantaneous states without loops exists in the system, then $\mathbf{P}=\mathbf{RE}^{c_{\max}}$ where c_{\max} is the maximal number of consecutive instantaneous states among all queues of consecutive instantaneous states. However, since $\mathbf{RE}^{c_{\max}}=\mathbf{RE}^{c_{\max}+a}$ for any integer $a > 0$, using a higher number will not generate a different \mathbf{P} matrix due to the definitions of \mathbf{R} and \mathbf{E} .

Thus, in general we use $\mathbf{P}=\mathbf{RE}^c$ where c is the number of instantaneous states in the model.

In the diabetic heart disease model in Figure 1, $c=1$ and therefore:

$$\mathbf{P}=\mathbf{RE}^1 = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & 0 & q_{02}q_{23} & q_{02}(1 - q_{23}) \\ 0 & 1 - q_{12} - q_{14} & 0 & q_{12}q_{23} & q_{12}(1 - q_{23}) + q_{14} \\ 0 & 0 & 0 & q_{23} & 1 - q_{23} \\ 0 & 0 & 0 & 1 - q_{32}(1 - q_{23}) - q_{34} & q_{32}(1 - q_{23}) + q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{No CVD} \\ \text{Angina} \\ \text{MI} \\ \text{Hx MI} \\ \text{Death} \end{array}$$

In this \mathbf{P} matrix, all elements in the column corresponding to the instantaneous state (state 2) are zero. This reflects the fact that at each transition time, no subject can be observed to stay in the instantaneous state.

2.2. Handling secondary data

The above matrix is appropriate when the study data agree perfectly with the theoretical model. When secondary data are used, the distinction between the study design and the theoretical model is critical. We follow the assumptions of [3] (the data are independent, unbiased, informative, and considered only the first occurrence of an event). Note that we use the term data to indicate summary statistics reported in the clinical literature, rather than the study's raw data.

Without losing generality, in this paper, a single study will provide information on a single transition (i.e., only one start state and only one end state) and can provide cumulative progression counts observed at different time points. For ease of explanation, we first only consider studies that only provide cumulative progression counts at one time point.

Let,

$(a_{(k)}, z_{(k)})$ denote the pair of indices for the start state and end state for the k^{th} study,

$t_{(k)}$ denote the units of time observed in the k^{th} study (study length),

$m_{(k)}$ denote the number of subjects in the k^{th} study (start population),

$x_{(k)}$ denote the number of subjects in the k^{th} study that have progressed from state $a_{(k)}$ to state $z_{(k)}$ by time $t_{(k)}$, and

$\pi_{k\ ij}(t_{(k)})$ denote the cumulative transition probability from state i to state j by time $t_{(k)}$ for the k^{th} study.

Similar to [3], from the assumption of independent studies, the likelihood can be expressed as

$$L \propto \prod_k [\pi_{(k)az}(t_{(k)})]^{x_{(k)}} [1 - \pi_{(k)az}(t_{(k)})]^{m_{(k)} - x_{(k)}}$$

Then we define $\mathbf{P}_{(k)}$ such that $\mathbf{P}_{(k)} \{ \begin{smallmatrix} (k) & ij \end{smallmatrix} \} = \pi_{(k)ij}(1)$, which is the unit time transition probability for the k^{th} study. $\mathbf{P}_{(k)}^{t_{(k)}}$ generates the cumulative probability of transition and depends on both the structure of \mathbf{P} and the design of the clinical study.

To use the secondary data available in the literature, three classes of a study's deviation from the theoretical model must be considered: 1) when an instantaneous state is a terminal state in a study, 2) when a non-instantaneous state in the theoretical model is a terminal state in a study, 3) when states in the theoretical model are pooled in a study [3].

Designed absorption is appropriate when a study has an outcome in a state that is not a terminal state in the theoretical model. For the first two classes of deviations that deal with the terminal state, we extend the approach of [3] by applying their concept of designed absorption to \mathbf{E} and to \mathbf{P} .

Formally, we define the sinking operator ψ such that when applied to a matrix \mathbf{B} and terminal state z it generates the following sunk matrix:

$$\psi(\mathbf{B}, z) = \begin{cases} b_{ij} & i \neq z \\ 0 & i=z, i \neq j \\ 1 & i=j=z \end{cases}$$

This sunk matrix nullifies the row associated with the terminal state z and places a 1 on the diagonal. Applying this sinking operator to a probability matrix means 0 probability of leaving z and a 100% probability of staying in this terminal state.

For the first class of deviation (when an instantaneous state is the terminal state in a study), $\mathbf{P}_{(k)} = \mathbf{R}(\psi(\mathbf{E}, z_{(k)}))^c$. The second class of deviation from the theoretical model occurs when the study design has terminal states that are not instantaneous states, and $\mathbf{P}_{(k)} = \psi(\mathbf{R}\mathbf{E}^c, z_{(k)})$

The third class of deviation from the theoretical model occurs when the study design generates realizations from a grouped Markov chain (i.e., the study does not differentiate between several states). For example, a study might follow subjects who either had no CVD or only had experienced angina and report the total number of subjects who develops MI by a certain time period without distinguishing the two types of subjects at the beginning of the study. To accommodate these additional inconsistencies between the study design and theoretical model, we apply Isaman's study-specific matrix [3] that we denote here as $\mathbf{K}_{(k)}(\mathcal{W})$. Briefly, Isaman constructs a study-specific matrix, $\mathbf{K}_{(k)}(\mathcal{W})$, that uses prevalence estimates, \mathcal{W} , to pool probabilities when study data are drawn from a mixture distribution of states (or subjects) in a theoretical model. Then $\mathbf{K}_{(k)}(\mathcal{W})$ is used to transform the transition matrix of the theoretical model, \mathbf{P} , into a study-specific transition matrix, \mathbf{P}_k , which is correct for the design of the k^{th} study. More specifically, $\mathbf{P}_{(k)} = \mathbf{K}_{(k)}(\mathcal{W}) \mathbf{P} \mathbf{K}_{(k)}(1)$. For more details, please refer to [3].

For a study with both deviations type one and type three, $\mathbf{P}_{(k)} = \mathbf{K}_{(k)} (\mathcal{W}) \Psi(E, z_{(k)})^c \mathbf{K}_{(k)}(1)$.
 For a study with both deviations type one and two, $\mathbf{P}_{(k)} = \Psi \mathbf{K}_{(k)} (\mathcal{W}) \mathbf{R} E^c \mathbf{K}_{(k)}(1), z_{(k)}$.

Since the method for dealing with the third type of deviation has been demonstrated in [3], it is not used or demonstrated in this paper. Unless stated otherwise, $\mathbf{K}_{(k)} (\mathcal{W}) = \mathbf{I}$.

When the data are longitudinal counts over time, the likelihood contributed by the k^{th} study is

$$L_{(k)} \propto [\pi_{(k)az}(t_{(k)1})]^{x_{(k)1}} \prod_{h=2}^{s-1} [\pi_{(k)az}(t_{(k)h}) - \pi_{(k)az}(t_{(k)h-1})]^{(x_{(k)h} - x_{(k)h-1})} [1 - \pi_{(k)az}(t_{(k)s-1})]^{m_k - x_{(k)s}}$$

where $t_{(k)h}$ are the measurement intervals and $0 < t_{(k)1} < \dots < t_{(k)s} = t_{(k)}$, and $x_{(k)h}$ is the cumulative number of subjects progressing to the end state by time $t_{(k)h}$. $t_{(k)}$ is the total length of study.

Our goal is to estimate parameters in the theoretical model, i.e., q_{ij} 's. Since $\pi_{kij}(t_{(k)})$ are derived from the \mathbf{P} matrix, which is in turn derived from the \mathbf{E} and \mathbf{R} matrices, the likelihood function (1) is a function of q_{ij} 's. Therefore, maximum likelihood parameter estimates for q_{ij} 's can be obtained by maximizing function (1). Variance of parameter estimates can be calculated from the empirical information matrix. The resulting estimates will have the usual asymptotic properties of maximum likelihood estimators (MLEs).

3. Simulation

We performed a number of simulations to investigate the finite sample property of our estimators using the following example (Figure 2), in which state 1 is an instantaneous state represented by a rhombus.

In each simulation, all studies were constructed under a known theoretical model with $q_{01}=0.2, q_{12}=0.3, q_{21}=0.4$.

The four types of direct data have the four following paths:

Study1: (0, 1)

Study2: (1, 2)

Study3: (1, 3)

Study4: (2, 1)

Since this model allows transition from state 2 to state 1, one can study the indirect path that potentially passes state 1 multiple times. For simplicity, here we only consider three indirect paths that pass the instantaneous state just once, and they are:

Study5: (0, 2)

Study6: (0, 3)

Study7: (2, 3)

According to the model presented in Figure 2

$$\mathbf{R} = \begin{bmatrix} 1 - q_{01} & q_{01} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & q_{21} & 1 - q_{21} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & q_{12} & q_{13} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For study 3, 6, and 7, since state 3 is an outcome state of both the study and the model, no modification is required for \mathbf{P} , and therefore:

$$\mathbf{P}_{(3)} = \mathbf{P}_{(6)} = \mathbf{P}_{(7)} = \mathbf{P} = \mathbf{R}\mathbf{E}^1 = \begin{bmatrix} 1 - q_{01} & 0 & q_{01}q_{12} & q_{01}q_{13} \\ 0 & 0 & q_{12} & q_{13} \\ 0 & 0 & 1 - q_{21}q_{13} & q_{21}q_{13} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For study 1 and 4, note that the study design ignores any states following state 1. Thus, in the study-specific matrix $\mathbf{P}_{(1)}$ and $\mathbf{P}_{(4)}$, the outcome (state 1) should be treated as a sink rather than an instantaneous state. Therefore the study-specific transition matrix is

$$\mathbf{P}_{(1)} = \mathbf{P}_{(4)} = \mathbf{R}\psi(\mathbf{E}, 1) = \mathbf{R}\mathbf{I} = \begin{bmatrix} 1 - q_{01} & q_{01} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & q_{21} & 1 - q_{21} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For study 2 and 5, designed absorption is manifested by placing a sink for state 2 when calculating $\mathbf{P}_{(2)}$ and $\mathbf{P}_{(5)}$ such that:

$$\mathbf{P}_{(2)} = \mathbf{P}_{(5)} = \psi(\mathbf{R}\mathbf{E}, 2) = \begin{bmatrix} 1 - q_{01} & 0 & q_{01}q_{12} & q_{01}q_{13} \\ 0 & 0 & q_{12} & q_{13} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that $q_{12} + q_{13} = 1$ as state 1 is an instantaneous state.

Since a subject who passes state 1 must immediately transit to either state 2 or state 3, Study 2 and 3 provide equivalent information. Same statement applies to Study 5 and 6. Therefore, in the simulations, we only use studies 1, 2, 4, 5, and 7.

In the first set of the simulation, we vary the number of studies per transition and number of subjects per study. We ran 1000 replications with 500 or 1000 subjects for each situation.

For, Study 1, 4, 5, and 7, each observation was generated by simulating the number of progressions among the subjects over a 2-year study. For Study 2, since it only counts events emanating from the instantaneous state 1 to the next directly linked state 2, and theoretically, the study takes no time, we arbitrarily assume the study length is 1 year.

Then, using the matrix multiplication for transition matrices in a discrete-time Markov chain based on [9] shown above, we have the following contributions of each type of studies to the log likelihood

$$\log L_{(1)=x_{(1)}} = x_{(1)} \log((1 - q_{01})q_{01} + q_{01}) + (m_{(1)} - x_{(1)}) \log(1 - (1 - q_{01})q_{01} - q_{01})$$

$$\log L_{(2)=x_{(2)}} = x_{(2)} \log(q_{12}) + (m_{(2)} - x_{(2)}) \log(1 - q_{12})$$

$$\log L_{(4)=x_{(4)}} = x_{(4)} \log(2q_{21} - q_{21}^2) + (m_{(4)} - x_{(4)}) \log(1 - 2q_{21} + q_{21}^2)$$

$$\log L_{(5)=x_{(5)}} = x_{(5)} \log((1 - q_{01})q_{01}q_{12} + q_{01}q_{12}) + (m_{(5)} - x_{(5)}) \log(1 - (1 - q_{01})q_{01}q_{12} - q_{01}q_{12})$$

$$\log L_{(7)=x_{(7)}} = x_{(7)} \log((1 - q_{21}q_{13})q_{21}q_{13} + q_{21}q_{13}) + (m_{(7)} - x_{(7)}) \log(1 - (1 - q_{21}q_{13})q_{21}q_{13} - q_{21}q_{13})$$

Table 1 displays the results of our first set of simulations. The table reports the number of studies for study types 1, 2, 4, 5, and, 7, respectively, as ratios (e.g., 1:1:1:1:1).

Since the usual properties for the maximum likelihood estimators apply, our estimators are consistent and asymptotically normal. For the sample size 500 and 1000, under all scenarios, the estimated standard error is very close to the empirical standard error on average.

Comparing (1:1:1:0:0) to (1:1:1:1:1), one can see that additional information from augmentary data decreases the standard error of \hat{q}_{01} , \hat{q}_{12} , and \hat{q}_{21} .

In the two scenarios where direct data is not available for one of the primary transition, using our method, we are able to estimate the transition probabilities through using augmentary data. In both of these two scenarios, comparing to (1:1:1:1:1) the standard error (SE) has increased quite a bit for the estimated transition probability with missing direct data.

In the (0:1:1:9:1) scenario, with 9 augmentary data that provide information for transition 0 to 1, SE decreased to almost half the size as in (0:1:1:1:1).

In the last scenario, with sample size increased to 1000 from 500, relative bias becomes smaller and SE decreased to about two thirds as in (0:1:1:1:1).

In addition, in order to show the consequence of mistakenly modeling an instantaneous state as a non-instantaneous state, we estimate model parameter in the scenario (1:1:1:1:1) by treating state 1 as a non-instantaneous state. We ran 1000 replications with 500 subjects for this simulation. Table 2 shows the result.

The result shows that when treating state 1 wrongly as a non-instantaneous state, the model does not allow a subject to transit from state 0 to state 2 or state 3 in one unit of time. Therefore, as one might expect, q_{01} and q_{21} are inflated, and q_{12} is deflated to compensate for this lower speed of transition. In other words, to achieve the transition rates indicated by studies with an event, a model with a delay has to increase flow to compensate that delay.

4. Modeling Heart Disease in Diabetes

We apply the Lemonade Method and the extension presented in this paper to the heart disease subprocess in the Michigan Model of Diabetes [2]. Figure 3 shows selected clinical data (indirect or direct) available for estimation together with the theoretical model shown in Figure 1. Note that in this model there is no transition from no CVD to death because the sub-process considers only CVD and such a transition will imply death from another cause.

Clinical data were extracted from the published medical literature based on the quality of the study design. Gray dashed lines in Figure 3 marked with capital letters are used to depict information provided by clinical studies. Studies were nominated for inclusion in the modeling effort whether or not they directly estimated parameters of interest. For example, Study D [10] investigated the transition from angina to death, ignoring the intermediary states. In contrast, Study F2 [11] directly estimates q_{32} . Details about these clinical data are provided in Table 3. These data are presented as cumulative counts of progression for the duration of years indicated. Often cumulative counts are the only data available in the literature. Estimated annualized rates for single cumulative counts were calculated for comparison between studies and for comparison with our final estimates.

The formula for calculating the annualized estimate is $1 - \left(1 - \frac{x(k)}{m(k)}\right)^{1/t(k)}$, which converts the cumulative probability of progression by the end of the study, $\frac{x(k)}{m(k)}$, as observed by the study after $t(k)$ years, and translates it into the probability of progression in one year that can now be compared to q_{ij} .

Note that in Study E [13], both in and out of hospital deaths were included. The annualized column was calculated using the stated equation as reference.

The UKPDS data denoted as Study B2 [8] are unique in that this study provides a risk equation (rather than incident counts) to summarize its data. The approach presented in this paper does not allow the use of risk equations, so we first generated the expected survival for male and female separately in the UKPDS population. To calculate these two expected survival counts, we set other variables in the UKPDS risk engine as the following: 7.6% Afro-Caribbean men (8.1% for women), 34% smokers among men (25% for women) with HbA1c, systolic blood pressure, and lipid ratio equal to the mean in the UKPDS population for each subpopulation of men and women. Then summation of resultant expected cumulative counts for men and women are used and presented in Table 3.

We apply the approach described above to these data to estimate the transition probabilities in our theoretical model. The MI state was treated as an instantaneous state. The relevant transition matrices \mathbf{E} and \mathbf{R} are constructed as presented in sections 2.1 and in the appendix. The study-specific partial likelihoods were generated using designed absorption when appropriate. The full likelihood function was maximized to obtain estimates for transition probabilities. For a detailed calculation for each study, see the appendix.

In addition, we also estimated the transition probabilities using a model that is exactly the same as the one shown in Figures 1 and 3, except that MI is treated as a regular state. Table 4 presents and compares parameter estimates using the model with MI as an instantaneous state and the model with MI as a regular state.

5. Discussion

In this paper, we have extended the Lemonade Method (an estimation method for discrete-state discrete-time model using indirect observation), to accommodate instantaneous state. The Lemonade Method provides two benefits over the conventional methods: It allows pooling data from multiple studies while reducing the variance of our estimates in the usual fashion. In addition, through the use of augmentary data, it allows one to estimate transition probabilities when no primary data are available.

Under the framework of discrete-state discrete-time models, including instantaneous states allows one to better model transitions that happen much faster than the rest of the transitions in the model. Consequently, this extension offers a better model for cost-effectiveness study when an instantaneous state exists. Using diabetic heart disease as an example, when MI is wrongly modeled as a regular state, a person who develops MI is not allowed to progress to death or survival of MI in the same year, therefore, the annual cost will only include the incident cost of MI, and the cost for death or survival of MI will be postponed to the next year. This leads to overall under-estimated costs for each year. When MI is modeled as an instantaneous state, it allows us to better estimate the cost by correctly weighting the occurrence of an MI separately from the annual costs of surviving an MI, or dying from MI.

Our method can be viewed as an extension of meta-analysis to multi-state models with more than a single outcome of interest. However, in contrast to meta-analysis, our method is focused on the natural history of disease rather than testing some clinical difference. As such, the published data used in our analyses are less prone to publication bias. In these “natural history” models, data are typically drawn from population-based studies (such as Ulvenstam [14]), national registries (such as The Surveillance, Epidemiology, and End Results, SEER), or from the control arm of clinical studies (such as the UKPDS).

Due to the discrete-time constraint, our model is limited to the assumption of no more than one instantaneous event in fixed time period, e.g., in our application of diabetic heart disease, it was necessary to assume no more than one MI in one year. In reality, multiple MI events can occur within a year. Another limitation of our method is our computational difficulty to accommodate the numerous covariates in the UKPDS risk engine and other studies providing covariates. The UKPDS provides a rich understanding of the contribution provided by continuous covariates such as age. Under our framework, only incident counts can be used as data, and the rich UKPDS data must be collapsed into a small table of counts, and with this approach, computational complexity will dramatically increase when adding many covariates. Isaman and colleagues [16] presented an alternative method to incorporate covariate information in published regression models such as in UKPDS under the discrete-time model frame work. Nonetheless, further extensions are possible. Continuous-time Markov chain models are a promising direction for future work direction which can potentially allow us to overcome these limitations of the current method.

Despite these limitations, we have demonstrated the ability to develop and test a clinical model using data that were previously unavailable to diabetes modelers. This provides us with an opportunity to use the accumulated knowledge in the field to answer clinical questions and provide a guide for long-term outcomes of a chronic disease such as diabetes.

The software implementation of the Lemonade Method is available under GPL license either as Matlab code or as Python code and can be freely downloaded from the project web site (<http://www.med.umich.edu/mdrtc/cores/DiseaseModel>) [17]. This software also has simulation capabilities described in [18].

Acknowledgments

We wish to thank Dr. Michael Brandle for his effort invested in developing the clinical model which motivated this research and extracting the estimates from the medical literature. The authors also wish to thank William H. Herman and Morton B. Brown for supervising this project.

This research was supported by the National Institutes of Health (NIH) Chronic Disease Modeling for Clinical Research Innovations R21-DK075077. Additional support provided by the Biostatistics Core of the Michigan Diabetes Research and Training Center (MDRTC) P60-DK20572.

Appendix

Appendix – Calculation Details for the Diabetic Heart Disease Example

This appendix presents more details regarding calculations of the diabetic heart disease model that are described in section 4.

As presented in table 4, the estimation process was conducted once for a model where MI is represented as an instantaneous state and once for the same model where MI is a regular state. The differences in results can be explained when examining the probability matrices generated for each study in each of the model variations. These matrices are the base for creating the likelihood expressions that are later calculated.

The main change between the models can be seen from the **E** and **R** matrices.

When MI is modeled as an instantaneous state, the matrices are:

$$\mathbf{R} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 - q_{12} - q_{14} & q_{12} & 0 & q_{14} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

When MI is modeled as a regular state these matrices are:

$$\mathbf{R} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 - q_{12} - q_{14} & q_{12} & 0 & q_{14} \\ 0 & 0 & 1 - q_{23} - q_{24} & q_{23} & q_{24} \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{E} = \mathbf{I}$$

Note that $q_{24} = 1 - q_{23}$.

These matrices are used to construct a probability matrix for each study. The probability matrix includes sinking according to the study end state. The probability matrices for one year are therefore:

Study A

When MI is modeled as an instantaneous state:

$$\mathbf{P} = \begin{bmatrix} 1 - q_{01} - q_{02}q_{23} - q_{02}q_{24} & q_{01} & 0 & q_{02}q_{23} & q_{02}q_{24} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 - q_{23} - q_{24} & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 - q_{32}q_{24} - q_{34} & q_{34} + q_{32}q_{24} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

When MI is modeled as a regular state:

$$\mathbf{P} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 - q_{23} - q_{24} & q_{23} & q_{24} \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Studies B1, B2, C, F1 and F2

Whether modeling MI as an instantaneous state or a regular state does not affect the \mathbf{P} matrix for these studies:

$$\mathbf{P} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 - q_{12} - q_{14} & q_{12} & 0 & q_{14} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Studies D, E, and G

When MI is modeled as an instantaneous state:

$$\mathbf{P} = \begin{bmatrix} 1 - q_{01} - q_{02}q_{23} - q_{02}q_{24} & q_{01} & 0 & q_{02}q_{23} & q_{02}q_{24} \\ 0 & 1 - q_{12}q_{23} - q_{12}q_{24} - q_{14} & 0 & q_{12}q_{23} & q_{12}q_{24} + q_{14} \\ 0 & 0 & 1 - q_{23} - q_{24} & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 - q_{32}q_{24} - q_{34} & q_{34} + q_{32}q_{24} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

When MI is modeled as a regular state:

$$\mathbf{P} = \begin{bmatrix} 1 - q_{01} - q_{02} & q_{01} & q_{02} & 0 & 0 \\ 0 & 1 - q_{12} - q_{14} & q_{12} & 0 & q_{14} \\ 0 & 0 & 1 - q_{23} - q_{24} & q_{23} & q_{24} \\ 0 & 0 & q_{32} & 1 - q_{32} - q_{34} & q_{34} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{R}$$

The differences between the results is thus driven by information provided in Studies A, D,E, and G. Curiously enough, the studies that directly end in an instantaneous state B1,B2,C,F1,F2 have a similar formulation whether the model uses MI as an instantaneous state or whether MI is modeled as a regular state. This is due to the fact that the studies view the state as a sink state that accumulates incidences. The model assumption of an MI being

instantaneous is therefore overridden by the study and due to this reason the instantaneous state has no influence on calculation results for these studies.

References

1. Herman WH. Diabetes modeling. *Diabetes Care*. 2003; 26:3182–3183. [PubMed: 14578260]
2. Zhou H, Isaman DJM, Messinger S, Brown MB, Klein R, Brandle M, Herman WH. A computer simulation model of diabetes progression, quality of life, and cost. *Diabetes Care*. Dec; 2005 28(12): 2856–63. [PubMed: 16306545]
3. Isaman DJM, Herman WH, Brown MB. A discrete-state and discrete-time model using indirect estimates. *Statistics in Medicine*. 2006; 25:1035–1049. DOI:10.1002/sim.2241. [PubMed: 16416413]
4. Ramsey SD, McIntosh M, Etzioni R, Urban N. Simulation modeling of outcomes and cost effectiveness. *Hematol Oncol Clin North Am*. 2000; 14:925–398. [PubMed: 10949781]
5. Schlessinger L, Eddy DM. Archimedes: a new model for simulating health care systems: the mathematical formulation. *J Biomedical Informatics*. 2002; 35:37–50. doi:10.1016/S1532-0464(02)00006-0.
6. Hoerger TJ, Hicks KA, Bethke AD. A Markov Model of Disease Progression and Cost-Effectiveness for Type 2 Diabetes. 2004 Technical Report, RTI Health, Social, and Economics Research, Funded by Centers for Disease Control and Prevention.
7. Mueller E, Maxion-Bergemann S, Gulytaev D, Walzer S, Freemantle N, Mathieu C, Bolinder B, Gerber R, Kvasz M, Bergemann R. Development and Validation of the Economic Assessment of Glycemic Control and Long-Term Effects of Diabetes (EAGLE) Model. *Diabetes Technology & Therapeutics*. 2006; 8(2):219–236. [PubMed: 16734551]
8. Stevens R, Kothari V, Adler A, Stratton I. The UKPDS risk engine: A model for the risk of coronary heart disease in type II diabetes UKPDS 56. *Clin Science*. 2001; 101:671–679.
9. Billingsley, P. *Statistical inference for Markov processes*. University of Chicago Press; 1961.
10. Malmberg K, Yusuf S, Gerstein H, Brown J, Zhao F, Hunt D, Piegas L, Calvin J, Keltai M, Budaj A. Impact of diabetes on long-term prognosis in patients with unstable angina and non-Q-wave myocardial infarction: results of the OASIS (Organization to Assess Strategies for Ischemic Syndromes) Registry. *Circulation*. 2000; 102(9):1014–1019. [PubMed: 10961966]
11. Haffner S, Lehto S, Ronnema T, Pyorala K, Laasko M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med*. 1998; 339:229–234. [PubMed: 9673301]
12. UKPDS: UK Prospective Diabetes Study UKPDS Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes UKPDS 33. *Lancet*. 1998; 352:837–853. doi:10.1016/S0140-6736(98)07019-6. [PubMed: 9742976]
13. Miettinen H, Lehto S, Salomaa V, Mahonen M, Niemela M, Haffner S, Pyörälä K, Tuomilehto J. Impact of diabetes on mortality after the first myocardial infarction. *Diabetes care*. 1998; 21:69–75. [PubMed: 9538972]
14. Ulvenstam G, Aberg A, Bergstrand R, Johansson S, Pennert K, Vedin A, Wilhelmsson L, Wilhelmsson C. Long term prognosis after myocardial infarction in men with diabetes. *Diabetes*. 1985; 34:787–792. [PubMed: 4018416]
15. Lowel H, Koenig W, Engel S, Hormann A, Keil U. Impact of diabetes on survival after myocardial infarction. *Diabetologia*. 2000; 43:218–226. doi: 10.1007/s001250050032. [PubMed: 10753044]
16. Isaman DJM, Barhak J, Ye W. Indirect Estimation of a Discrete-State Discrete-time model using Secondary Data Analysis of Regression Data. *Statistics in Medicine*. 2009; 28(16):2095–2115. DOI: 10.1002/sim.3599. [PubMed: 19455575]
17. Michigan Diabetes Research and Training Center. Disease Modeling Software for Clinical Research (Online). [accessed on 22 July 2009]. Online: <http://www.med.umich.edu/mdrtc/cores/DiseaseModel>
18. Barhak J, Isaman DJM, Ye W, Lee D. Chronic disease modeling and simulation software. *Journal of Biomedical Informatics*. 2010 Article in press. doi:10.1016/j.jbi.2010.06.003.

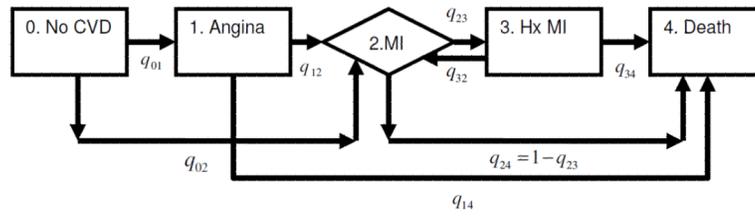


FIGURE 1.
A Model of Diabetic Heart Disease

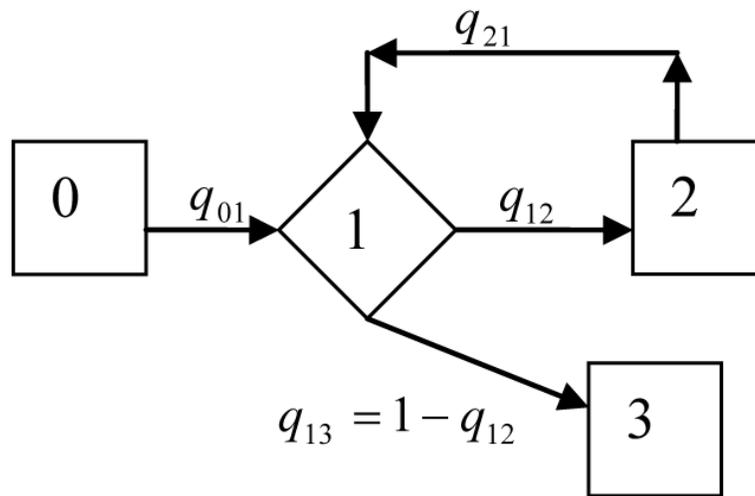


FIGURE 2. Diagram of a theoretical model for a simple example: Rhombus indicates instantaneous state.

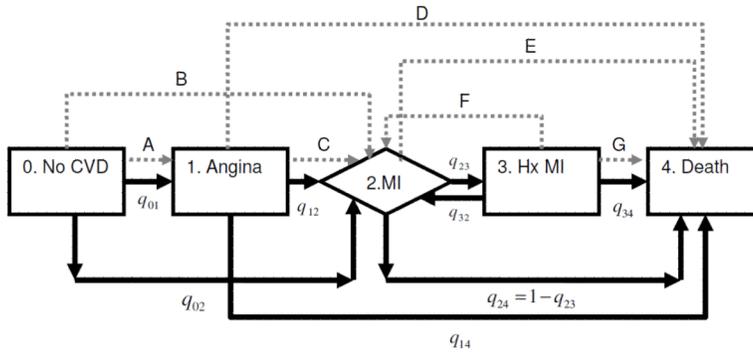


FIGURE 3. Full estimation model definitions (denoted by dark arrows and probabilities) and study transitions (denoted by dotted gray and capital letters).

TABLE 1

Simulation results with 1000 replications of a 4-node model

| Feature | Studies | Sample size | Parameter | Average estimate | Relative Bias(%) | Empirical SE | Average estimated SE | 95%CI coverage rate (%) |
|-------------------|-------------|-------------|-----------------|------------------|------------------|--------------|----------------------|-------------------------|
| No augmentary | (1:1:1:0:0) | 500 | q ₀₁ | 0.1998 | -0.1 | 0.0141 | 0.0134 | 93.5 |
| | | | q ₁₂ | 0.3001 | 0.33 | 0.0208 | 0.0205 | 95.3 |
| | | | q ₂₁ | 0.4001 | 0.25 | 0.0176 | 0.0179 | 95.2 |
| | (1:1:1:1:1) | 500 | q ₀₁ | 0.2002 | 0.10 | 0.0129 | 0.0124 | 93.5 |
| | | | q ₁₂ | 0.2996 | -0.13 | 0.0169 | 0.0173 | 94.8 |
| | | | q ₂₁ | 0.4003 | 0.08 | 0.0138 | 0.0145 | 96.5 |
| No primary 1 to 2 | (1:0:1:1:1) | 500 | q ₀₁ | 0.2002 | 0.10 | 0.0131 | 0.0129 | 95.2 |
| | | | q ₁₂ | 0.3002 | 0.06 | 0.0330 | 0.0323 | 94.5 |
| | | | q ₂₁ | 0.4008 | 0.20 | 0.0162 | 0.0157 | 94.2 |
| No primary 0 to 1 | (0:1:1:1:1) | 500 | q ₀₁ | 0.2024 | 1.20 | 0.0330 | 0.0325 | 94.8 |
| | | | q ₁₂ | 0.2990 | -0.33 | 0.0188 | 0.0189 | 95.3 |
| | | | q ₂₁ | 0.3998 | -0.05 | 0.0145 | 0.0146 | 95.4 |
| No primary 0 to 1 | (0:1:1:9:1) | 500 | q ₀₁ | 0.2008 | 0.40 | 0.0174 | 0.0173 | 95.0 |
| | | | q ₁₂ | 0.3003 | 0.10 | 0.0189 | 0.0189 | 94.7 |
| | | | q ₂₁ | 0.4007 | 0.18 | 0.0145 | 0.0146 | 96.1 |
| No primary 0 to 1 | (0:1:1:1:1) | 1000 | q ₀₁ | 0.2007 | 0.35 | 0.0230 | 0.0229 | 94.3 |
| | | | q ₁₂ | 0.3003 | 0.10 | 0.0132 | 0.0134 | 95.6 |
| | | | q ₂₁ | 0.4000 | 0.00 | 0.0103 | 0.0103 | 94.9 |

Table 2

Comparative solution for scenario (1:1:1:1:1): treating state 1 as an instantaneous state and a non-instantaneous state

| Method | Parameter | average | Relative Bias (%) | Empirical SE | Average estimated SE | 95% CI coverage rate |
|---|-----------------|---------|-------------------|--------------|----------------------|----------------------|
| Treating state 1 as instantaneous state | q ₀₁ | 0.1993 | -0.35 | 0.0123 | 0.0124 | 94.9 |
| | q ₁₂ | 0.3004 | 0.13 | 0.0174 | 0.0173 | 95.1 |
| | q ₂₁ | 0.4010 | 0.25 | 0.0143 | 0.0145 | 96.1 |
| Treating state 1 as non-instantaneous state | q ₀₁ | 0.2210 | 10.5 | 0.0129 | 0.0131 | 64.0 |
| | q ₁₂ | 0.2776 | -7.5 | 0.0159 | 0.0162 | 70.9 |
| | q ₂₁ | 0.4759 | 19.0 | 0.0156 | 0.0155 | 0.3 |

TABLE 3

Study data used for estimation of model probabilities

| Transition | Number of Subjects $m_{(k)}$ | Cumulative Count $x_{(k)}$ | Years $t_{(k)}$ | Annualized $\frac{x_{(k)}}{1 - \left(1 - \frac{x_{(k)}}{m_{(k)}}\right)^{1/t_{(k)}}}$ | Reference |
|------------|------------------------------|----------------------------|-----------------|---|----------------|
| A: 0 to 1 | 1138 | 72 | 10 | 0.0065 | UKPDS [12] |
| B1: 0 to 2 | 890 | 180 | 7 | 0.032 | Haffner [11] |
| B2: 0 to 2 | 4540 | 70, 150, 242, 345, 462 | 2,4,6,8,10 | | Stevens [8] |
| C: 1 to 2 | 569 | 61 | 2 | 0.055 | Malmberg [10] |
| D: 1 to 4 | 569 | 53 | 2 | 0.048 | Malmberg [10] |
| E: 2 to 4 | 620 | 268 | 1 | 0.4323 | Miettinen [13] |
| F1: 3 to 2 | 73 | 13, 20, 34 | 1,2,5 | | Ulvénstam [14] |
| F2: 3 to 2 | 169 | 76 | 7 | 0.082 | Haffner [11] |
| G: 3 to 4 | 403 | 137 | 5 | 0.0787 | Lowel [15] |

TABLE 4

Comparing estimation results of diabetic heart disease models

| | Model 1: With MI as an instantaneous state | | Model 2: With MI as a regular state (non-instantaneous state) | |
|----------|--|----------------|---|----------------|
| | MLE of transition prob. | Standard Error | MLE of transition prob. | Standard Error |
| q_{01} | 0.0070 | 0.0008 | 0.0070 | 0.0008 |
| q_{02} | 0.0119 | 0.0006 | 0.0119 | 0.0006 |
| q_{12} | 0.0569 | 0.0069 | 0.0573 | 0.0070 |
| q_{14} | 0.0225 | 0.0072 | 0.0358 | 0.0066 |
| q_{23} | 0.5686 | 0.0199 | 0.5685 | 0.0199 |
| q_{32} | 0.1032 | 0.0088 | 0.1078 | 0.0096 |
| q_{34} | 0.0362 | 0.0070 | 0.0476 | 0.0071 |