

Article

Data Analysis of the Risks of Type 2 Diabetes Mellitus Complications before Death Using a Data-Driven Modelling Approach: Methodologies and Challenges in Prolonged Diseases

Ming-Yen Lin ^{1,2}, Jia-Sin Liu ³, Tzu-Yang Huang ⁴, Ping-Hsun Wu ^{1,2} , Yi-Wen Chiu ^{1,2} , Yihuang Kang ⁵ , Chih-Cheng Hsu ⁶, Shang-Jyh Hwang ^{1,2,5,7} and Hsing Luh ^{4,*} 

- ¹ Division of Nephrology, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 807, Taiwan; mingyenlin3@gmail.com (M.-Y.L.); 970392@kmuh.org.tw (P.-H.W.); chiuyiwen@gmail.com (Y.-W.C.); sjhwang@kmu.edu.tw (S.-J.H.)
 - ² Department of Renal Care, College of Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
 - ³ Department of Public Health, College of Health Science, Kaohsiung Medical University, Kaohsiung 807, Taiwan; sgazn.tw@gmail.com
 - ⁴ Department of Mathematical Sciences, National Chengchi University, Taipei City 116, Taiwan; kobe24abc@gmail.com
 - ⁵ Department of Information and Management, National Sun Yat-Sen University, Kaohsiung 804, Taiwan; ykang@mis.nsysu.edu.tw
 - ⁶ Institute of Population Health Sciences, National Health Research Institutes, Miaoli County 350, Taiwan; cch@nhri.edu.tw
 - ⁷ Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung 807, Taiwan
- * Correspondence: sl@nccu.edu.tw



Citation: Lin, M.-Y.; Liu, J.-S.; Huang, T.-Y.; Wu, P.-H.; Chiu, Y.-W.; Kang, Y.; Hsu, C.-C.; Hwang, S.-J.; Luh, H. Data Analysis of the Risks of Type 2 Diabetes Mellitus Complications before Death Using a Data-Driven Modelling Approach: Methodologies and Challenges in Prolonged Diseases. *Information* **2021**, *12*, 326. <https://doi.org/10.3390/info12080326>

Academic Editor: Thomas T.H. Wan

Received: 6 April 2021

Accepted: 21 May 2021

Published: 12 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: (1) Background: A disease prediction model derived from real-world data is an important tool for managing type 2 diabetes mellitus (T2D). However, an appropriate prediction model for the Asian T2D population has not yet been developed. Hence, this study described construction details of the T2D Holistic Care model via estimating the probability of diabetes-related complications and the time-to-occurrence from a population-based database. (2) Methods: The model was based on the database of a Taiwan pay-for-performance reimbursement scheme for T2D between November 2002 and July 2017. A nonhomogeneous Markov model was applied to simulate multistate (7 main complications and death) transition probability after considering the sequential and repeated difficulties. (3) Results: The Markov model was constructed based on clinical care information from 163,452 patients with T2D, with a mean follow-up time of 5.5 years. After simulating a cohort of 100,000 hypothetical patients over a 10-year time horizon based on selected patient characteristics at baseline, a good predicted complication and mortality rates with a small range of absolute error (0.3–3.2%) were validated in the original cohort. Better and optimal predictabilities were further confirmed compared to the UKPDS Outcomes model and applied the model to other Asian populations, respectively. (4) Contribution: The study provides well-elucidated evidence to apply real-world data to the estimation of the occurrence and time point of major diabetes-related complications over a patient's lifetime. Further applications in health decision science are encouraged.

Keywords: type 2 diabetes; complication; discrete event simulation; electronic health record

1. Introduction

The incidence of type 2 diabetes (T2D) has reached epidemic proportions globally, with an average of one person dying of diabetes every 8 s in the world [1]. In South-East Asia regions, T2D has been estimated to have increased by 74% from 2019 to 2045, resulting in a great economic impact [1,2]. Inappropriate T2D care and management have directly caused permanent morbidities, including cardiovascular disease (CVD), kidney

disease, blindness, lower-limb amputations, and mortality. It has been estimated that T2D substantially contributed to approximately 32% of all CVD [3], 56% of new end-stage renal disease [4], and 1.6 million deaths in 2016 [5]. Hence, seeking optimal solutions to control T2D-associated complications is critical to relieve the burden of T2D.

Diabetes ranks fifth in the top ten causes of death among Taiwanese people, with diabetic patients frequently experiencing hypertension and hyperlipidaemia, which dramatically increase the risk of neuropathy, as well as microvascular and macrovascular complications [6]. Diabetes Health Education advises the treatment goals of glycated-heme control to be below 7%, blood pressure below 130/80 mmHg, and low-density cholesterol less than 100 mg/dL, to reduce the associated CVD burden. In Taiwan, there are approximately 2 million diabetic patients clinically at risk of diabetic complications. Thus, the Taiwan Diabetes Health Education Association has targeted nearly 6000 diabetes patients with glycosylated haemoglobin A1c (HbA1c) and blood pressure and cholesterol, the so-called ABC, to investigate indicators of complications of diabetes. Since 2002, patients who have met the treated target of ABC have shown a continuous growth trend [7]. The survey found that the total compliance rate in Taiwan patients with diabetes was 12.9% in 2016. Compared with the total compliance rate in 2011, it increased by nearly 60%. The above-mentioned information combining detailed prescription and diagnosis data. The National Health Insurance Research Database (NHIRD) is appropriately used to build a risk assessment model for T2D patients.

Previous studies have explored several algorithms in assisting diabetes prevention and management. Zou et al. analysed data from 68,994 healthy people and patients with diabetes, and the present healthcare recommendation system using machine learning helped to predict the risk of diabetes. [8]. Simon et al. presented a survey for predictive models built to leverage big data analytics to estimate the possibility of developing complications in patients with diabetes [9]. Yap et al. [10] introduced mobile apps collecting images of diabetic patients' feet for data training and the development of diabetic foot ulcers. Ellaham [11] argued that since intelligent technical reforms have produced better glycaemic control with reductions in fasting and postprandial glucose levels, glucose excursions, and glycosylated haemoglobin, AI will introduce a paradigm shift in diabetes care from conventional management strategies to building targeted data-driven precision care.

Since the publication of the first United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model [12], the use of simulation modelling in diabetes has increased. For example, a total of 39,043 person-years of data where the median follow-up time for patients in the ACCORD trial was 3.7 years supports the model fitting process for developing the BRAVO risk engine [13], with which the Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus (ASPEN) trial, the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) trial, and the Collaborative Atorvastatin Diabetes Study (CARDS) trial were compared. In summary, we present the external validation by simulation with comparable outcomes versus some reports excerpted from the Mount Hood Challenges [14,15]. Compared to previous models constructed based on western clinical trials or epidemiological studies, such as the Framingham and the UKPDS studies [12], developing the first model using long-term clinical care data collected from the Asian population would provide a more reliable estimate to help clinicians make better decisions regarding the treatment of diabetic patients.

We present a new version of the Taiwan Type 2 Diabetes Mellitus Holistic Care Model (T2DHoc), which could provide a well-elucidated estimation of the occurrence and time point of major diabetes-related complications over a patient's lifetime. The developed data-driven model could simulate the occurrence of competing diabetes-associated complications and all-cause death for T2D over a patient's lifetime. Since the participants had representative data, the model was also compared to other available prediction models and subject data from clinical trials. The internal validation of the T2DHoc over the 5-year follow-up was assessed by sensitivity analysis, and predictions from the original and

new models using a contemporary patient-level input dataset. The data-driven approach for the predictive model is presented by the revealing management of prevalent short and long-term complications in diabetes, where long-term is defined as at least 15 years follow-up.

Diabetes is one of the top priorities in medical science and health care management, and an abundance of data and information is available on these patients. Whether data stem from statistical models or complex pattern recognition models, they may be fused into predictive models that combine patient information and prognostic outcome results. Almost any statistical regression model can be used as a predictive model, but due to their transparent functionality, multiple logistic or similar linear regressions are often used for prediction model development [16]. However, a prediction model should take into account all important complication risks as a whole. To capture the whole disease as a system has been discussed in Tappenden et al. [17] and Esensoy and Carter [18].

Such knowledge could be used in clinical decision support, disease surveillance, and public health management to improve patient care. Predictive models often include multiple predictors (covariates) to estimate the probability, risk of a certain outcome or to classify that a certain outcome is present/absent (diagnostic prediction model) or will happen within a specific timeframe (prognostic prediction model) in an individual. Although extensive effort has been made with building these prediction models, there is a remarkable scarcity of impact studies due to data limitations. In Taiwan, the national diabetic care registry was established in 2001, so it is possible to collect enough information for long-term prediction with important covariates to appraise high risk events. Regarding the risk of diabetic complications, the primary events of this study were arteriosclerotic heart disease (ASHD), ischaemic heart disease (IHD), chronic heart failure (CHF), ischemic stroke (ISC), first-time renal failure (FESRD), retinopathy (EYE), amputation (Fin Foot), and death. The T2DHoc model was built upon the occurrence of complication risks after a diagnosis of diabetes. Figure 1 presents the logic structure with two layers, where the first layer after T2D demonstrates the first occurrence of complications and the second layer indicates the second complication. Therefore, the study aims to describe construction details of the T2DHoc model through estimating the probability of diabetes-related complications and the time-to-occurrence from a population-based database. In addition, we also validate this model against the results from Japan, Korean clinical studies, and the UKPDS Outcomes Model.

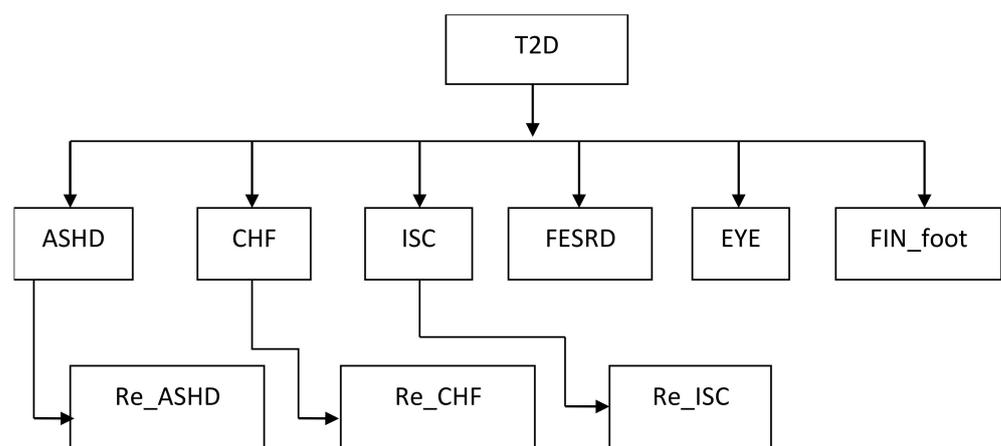


Figure 1. Logic structure of a prediction model. Abbreviation: T2D, type 2 diabetes; ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

2. Materials and Methods

2.1. Data Sources

The data for the model development were taken from the Taiwan T2DHoc using the Diabetic Pay-for-performance (P4P) Registry, a subset of the Taiwan NHIRD. The Taiwan National Health Insurance (NHI) programme initiated in 1995 is compulsory universal health insurance to provide holistic healthcare to over 99% of Taiwan's inhabitants. All contracted healthcare providers are mandated to upload medical claims; the database contains comprehensive information on insured subjects, including dates of clinical visits, diagnostic codes, details of prescriptions, medical procedures, and expenditure [2,3]. Taiwan's NHI Administration regularly performs reviews of the system to prevent waste, safeguard quality, and maintain public healthcare safety and quality [4]. The NHIRD has been opened to researchers for generating real-world population-based evidence to test clinical and epidemiological research hypotheses under governmental regulations. It includes approximately 2.4 million laboratory records, 60 million outpatient records, and 554,000 inpatient records.

In 2001, the NHI Administration launched a pay-for-performance programme to encourage healthcare providers to deliver high-quality care to patients with T2D. Patients with T2D and healthcare providers can voluntarily enrol in the program. During the early period of the program, it cultivated and certified T2D care providers and encouraged them to increase patient monitoring and follow-up [5]. To verify care quality, the laboratory data of patient care were uploaded under the T2D P4P scheme. This study was approved by the Institutional Review Boards of National Chengchi University and National Health Research Institutes, Taiwan (NCCU-REC-201603-I005 and EC1050505-E). All research procedures followed the directives of the Declaration of Helsinki. All identifying personal information was removed from the data files prior to analysis, so the review board waived the requirement for informed written consent.

In addition, the T2D cohort data were also verified through Kaohsiung Medical University Hospital Research Database (KMUHRD). Kaohsiung Medical University Hospital (KMUH) is a medical centre located in southern Taiwan, with around 1600 beds and 6000 clinical visits per day in 2015. The KMUHRD, which is managed by the KMUH Division of Medical Statistics and Bioinformatics, offers comprehensive data of approximately two million patients who attended KMUH from 2009 to 2015, with coverage on ambulatory care, hospital admissions, dental services, drug-dispensing records, and biochemical test results. For confidentiality and according to the Personal Information Protection Act, all personal identifiers were removed, and the authorised researchers only performed data linkage, processing, and statistical analyses with specified computers in an independent 24-hr monitored room using encrypted identifiers. In this study, all diagnoses were coded according to the International Classification of Diseases, 9th Revision and Clinical Modification (ICD-9-CM). Subjects with type 2 DM (ICD-9-CM codes 250.1–250.9) who were prescribed with hypoglycaemic agents and had an HbA1c level $\geq 6.5\%$ (≥ 48 mmol/mol) were included. The clinical definitions of diabetic complications of T2D patients are summarised in Appendix A.

2.2. Data-Driven Approaches

NHRID stores a huge amount of patient-specific data, including demographic information, diagnoses, laboratory tests, prescriptions, radiological images, and clinical notes, which can be extracted via neural networks and deep learning models. Data-driven thinking and methods which depart from traditional statistical tools or algorithms utilise advanced computational and mathematical systems to cope with the biomedical data to discover underlying patterns, disease phenotypes, unanticipated effects, etc. Dynamic intervention is often involved in healthcare systems. As the population needs change over time or interactions among clinical entities, it is difficult for decision-makers to plan healthcare services. Thus, with the capability of dealing with the variety and uncertainty in healthcare systems,

computer-based methods provide a more practical way to model diabetes management systems without costly and time-consuming experimentation [19–21].

Generally, there are two types of models for building a prediction model—parametric and nonparametric. Parametric models make assumptions regarding the underlying data distribution, whereas nonparametric models (and semiparametric models) make fewer or no assumptions about the underlying distribution. The model developed in this study is a combination of both, with respective time distributions and transition rates among complications. A total of 62 (13 + 49) different combinations of complications (endpoints) were predicted using the T2DHoc risk engine through a discrete-time event micro-simulation process under the corresponding time horizon of each possible diabetic complication.

Diabetes progression was characterised by changes in a patient’s disease status in terms of the number of complications that they developed. The number implies the occurrence of one event that may also preclude the development of another, thus, the major diabetes progression status of this study includes T2D-associated complications, all-cause of death, and staying in the same disease state while considering the competing risk events. Prognostic models with competing risks have been discussed in various studies [22–25]. To tackle the interaction among health states, we analysed the development of these competing risk events within a single mathematical analytic framework instead of treating each complication in isolation, which enables us to determine whether one event is more likely to occur than another. A tree-like structure of patients’ disease progression was set up as shown in Figure 2, consisting of 62 risk equations representing important pathways after diabetes has been diagnosed.

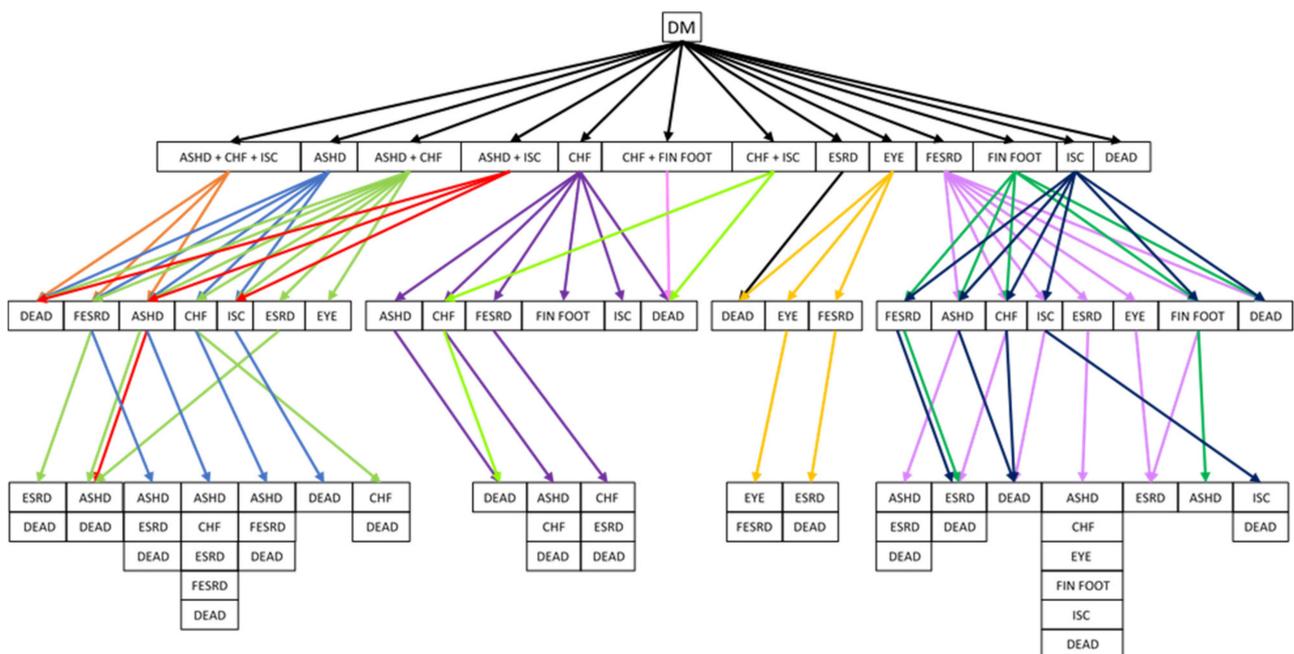
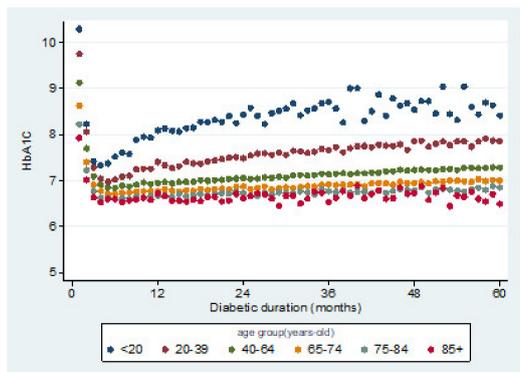


Figure 2. A tree structure of diabetes mellitus complications. Abbreviation: DM, diabetes mellitus; ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

The baseline characteristics of each trial have been reported and applied directly as the characteristic of simulation samples. Conventionally, Weibull distributions are useful to describe the survival time distribution, which was adopted from the corresponding literature [26]. The simulation assumes that all time transition variables are Weibull distributed between health states, but normal distributions are assumed for the standard error of some risk factors as simulation variates. HbA1c may change during the simulation, where the value function in Figure 3 is constructed based on clinical data. Table 1 gives

the distribution of values of key risk factors, including low-density lipoprotein (LDL), systolic blood pressure (SBP), urine, etc., which were assumed to reach the corresponding treatment target in the first year and used as input parameters for hypothetical patients. Some risk factors are assumed constant for each patient individually in the simulation.



$$x_{HbA1c}(a, t) = \begin{cases} \theta(a, t) + \varphi(a, t)x_{HbA1c}(a, t - 1) + \psi(a, t) \times t, & t < 60 \\ x_{HbA1c}(a, t - 1), & t \geq 60 \end{cases}$$

Figure 3. HbA1c values change with respect to time at different ages where the HbA1c value is a function of age ‘a’ and duration ‘t’, where θ , φ , and ψ are coefficients of this regression function.

Table 1. The baseline of the new diagnosis of diabetic subjects in the program of diabetic pay for performance in 2002–2016. Data represented as mean and SD, and count and proportion for continuous and categorical variables, respectively.

| Cohort | 2002–2007 | 2008–2010 | 2011–2014 | 2015–2016 | p Value |
|------------------------------------|---------------|---------------|---------------|---------------|---------|
| n | 48,026 | 18,369 | 54,250 | 42,807 | <0.001 |
| Age, mean (sd) | 55.3 (13.1) | 54.9 (13.3) | 55 (13.6) | 54.9 (13.6) | <0.001 |
| 00–19, n (%) | 375 (0.8) | 206 (1.1) | 707 (1.3) | 471 (1.1) | 0.02 |
| 20–39, n (%) | 4587 (9.6) | 1892 (10.3) | 5978 (11.0) | 5109 (11.9) | 0.03 |
| 40–64, n (%) | 31,003 (64.6) | 12,209 (66.5) | 35,359 (65.2) | 27,164 (63.5) | 0.019 |
| 65–74, n (%) | 8441 (17.6) | 2730 (14.9) | 8162 (15.0) | 7036 (16.4) | 0.03 |
| 75+, n (%) | 3573 (7.4) | 1315 (7.2) | 4008 (7.4) | 2988 (7.0) | 0.007 |
| Male, n (%) | 26,469 (55.1) | 10,027 (54.6) | 30,111 (55.5) | 24,127 (56.4) | 0.012 |
| Female, n (%) | 21,510 (44.8) | 8325 (45.3) | 24,103 (44.4) | 18,641 (43.5) | 0.012 |
| HbA1c % | 8.0 (20) | 7.8 (1.9) | 7.8 (2.0) | 7.7 (1.9) | <0.001 |
| SBP mmHg | 132.1 (19.1) | 130.8 (17.7) | 130.6 (17.1) | 131.1 (17.1) | <0.001 |
| DBP mmHg | 80.7 (13.1) | 79.6 (13.3) | 79.5 (15) | 79.3 (13.9) | <0.001 |
| Height, cm | 161.2 (8.8) | 161.8 (9.1) | 162.3 (9.3) | 162.5 (9.5) | <0.001 |
| Weight, Kg | 68.3 (13.5) | 69.4 (14.3) | 70.2 (15.4) | 70.5 (15.6) | <0.001 |
| BMI | 26.2 (4.3) | 26.4 (4.4) | 26.5 (4.7) | 26.5 (4.6) | <0.001 |
| Pulse (time) | 79.9 (14.8) | 80.5 (12.2) | 81.1 (12.4) | 81.8 (12.8) | <0.001 |
| Laboratory data | | | | | |
| LDL, mg/dL | 121.7 (98.8) | 115 (36.6) | 114.3 (37.5) | 113.8 (38.5) | <0.001 |
| BUN, mg/dL | 20.7 (24.8) | 14.9 (4.6) | 16.1 (6.6) | 14.4 (5.1) | <0.001 |
| HDL, mg/dL | 45.7 (12.8) | 46.8 (12.5) | 45.7 (12.1) | 44.9 (11.7) | 0.02 |
| Urine acid, mg/dL | 5.8 (1.8) | 5.8 (1.7) | 5.7 (1.6) | 5.6 (1.8) | <0.001 |
| Creatinine, mg/dL | 1.0 (0.5) | 0.9 (0.4) | 0.9 (0.4) | 0.9 (0.4) | <0.001 |
| Albumin, g/dL | 3.8 (1.6) | 4.4 (0.6) | 4.5 (4.9) | 5.1 (5.3) | <0.001 |
| Total cholesterol, mg/dL | 198.8 (48.2) | 188.1 (43.2) | 186.3 (45.5) | 186.6 (47.1) | <0.001 |
| Triglyceride, mg/dL | 178 (140.3) | 166.3 (126.7) | 171.4 (138) | 174.8 (143.3) | 0.009 |
| Urine | | | | | |
| microalbumin, mg/dL | 53.1 (216.2) | 48.2 (208.8) | 55.1 (234.3) | 59.4 (245.5) | <0.001 |
| >30, n (%) | 3205 (6.7) | 1623 (8.8) | 6845 (12.6) | 6990 (16.3) | 0.12 |
| >300 n (%), | 462 (1.0) | 240 (1.3) | 1021 (1.9) | 1093 (2.6) | 0.047 |
| >3000, n (%) | 15 (0.0) | 5 (0.0) | 42 (0.1) | 46 (0.1) | 0.012 |
| Urine protein, >0.05 g/24 h, | 6313 (13.1) | 2362 (12.9) | 7129 (13.1) | 1196 (2.8) | 0.15 |
| Urine protein >1 g/24 h, n (%) | 802 (1.7) | 505 (2.7) | 1908 (3.5) | 155 (0.4) | 0.09 |
| Urine protein >10 g/24 h, | 354 (0.7) | 131 (0.7) | 711 (1.3) | 75 (0.2) | 0.05 |
| Urine protein in g/24 h, mean (sd) | 34.7 (89.8) | 22.2 (107.3) | 37.9 (256.8) | 20.2 (63.8) | 0.014 |

Abbreviation: HbA1c, glycosylated haemoglobin A1c; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; LDL, low-density lipoprotein; BUN, blood urea nitrogen; HDL, high-density lipoprotein.

The prediction model is based on computer simulation techniques, a method of modelling the progression of T2D and predicting long-term disease outcomes. The development of the simulation model involved (1) re-estimating, over a longer duration of follow-up, the 13 risk equations for complications ASHD, ISC, CHF, Fin Foot, EYE, FESRD and those combinations, and (2) estimating the following risks after the first complication, which involves constructing sixty-two new equations (13 + 49) and some follow-up events.

2.3. Building Risk Equations

In this analysis, a proportional hazards Weibull regression model was used to model diabetes-related complications with a baseline hazard:

$$h_0(t) = \lambda \kappa t^{\kappa-1} \quad (1)$$

where κ is a shape parameter, and the scale parameter is $\lambda = \exp(\beta_0)$ which is expressed by the exponential intercept coefficient β_0 . According to the proportional hazards assumption, the hazard of an event at time t is:

$$h(t|X(t)) = h_0(t) \exp(X(t) * \beta) \quad (2)$$

where $X(t)$ is a vector of covariates and β is a vector of associated coefficients. The unknown parameters requiring estimation are λ , κ , β_0 and β . Some of these covariates (such as starting age and sex) remain constant as time elapses, while others potentially vary over time (such as HbA1c and SBP).

For each risk factor k , a regression learning algorithm by Bayesian approach finds $\beta_k(i, j)$ associated with X_k to represent a risk score during the time interval (t_0, t_1) bore by the patient of age a from health state i to state j (one complication). Here, for each pair (i, j) were required survival distributions assuming that no other event was possible. Then, a time was sampled for each event and the earliest time determines which event happens. This is implemented by taking other events as censored events and the other times are discarded. Namely, a risk score at time t between the time period $t_0 < t \leq t_1$ yields this patient a quantified value R_a that indicates the impact of this specified complication under health conditions $X_k, k = 1, 2, \dots, p$:

$$R_a(t, i, j) = \beta_0 + \sum_k \beta_k(i, j) X_k(a, t), \quad (3)$$

where $X_k(a, t)$, $\beta_k(i, j)$ are explanatory variables ($k = 1, \dots, p$) provided there are p cofactors to consider in this model. The Cox proportional hazard regression expression can be written as:

$$C_a(t, i, j) = \exp(R_a(t, i, j)) \quad (4)$$

According to the corresponding conditional assumptions, we suggest a risk equation that would be expressed by, $t_0 < t \leq t_1$:

$$r_a(t, i, j) = 1 - \exp\{[H(t_0) - H(t_1)]C_a(t, i, j)\} \quad (5)$$

provided that the accumulated hazard rate at time t is, $H(t)$. Model Equation (5) denotes the probability (risk) of a complication event during (t_0, t_1) for a patient with newly diagnosed diabetes, in the absence of death from causes other than this specific complication. Information retrieved from the collected data provides the essential details to the mechanism governing these transitions and serves as the major modelling tool in this study.

Predicted values of R_a were used in conjunction with the event equations to complete the simulations. This means that $H(t)$ may be estimated for a specific diabetes complication, such as CHF status equation or ISC from the diagnosis of diabetes [27]. The detailed description and methodology used for the prediction model are given in Appendix B.

Confidence intervals presented in this paper are based on a two-stage process of evaluation. Firstly, the original data used to fit the risk equations were bootstrapped, and

the risk equations refitted, and the coefficients recorded. Repeating this process many times generated a vector of coefficients that represented the parameter uncertainty in those coefficients, also accounting for the covariance between risk equations. As the model was applied to predict lifetime outcomes of P4P patients, in the second stage, we treated the predictions as imputations of missing values, in that we were predicting values that were not observed. Standard methods for combining the results of multiple imputations [28] were then employed, including a bias correction to adjust for the relatively small number of multiple imputations performed.

2.4. The Simulation

Based on the T2DHoc model, the computer simulation was executed by a cohort profile of hypothetical patients with the initial description of demographic characteristics and baseline risk factors as given in Table 1. In the simulation of diabetes disease, hypothetical patients were created based on random baseline risk factors. Patients have attributes in which each individual has a specific value for each characteristic set at the start of the simulation and may be updated as events occur, such as age increases, disease severity decreases, the number of risk events being incremented, etc. For example, HbA1c changes according to the function in Figure 3, or disease duration follows the time distribution given in Equation (2). The flowchart of the simulation procedure is shown in Figure 4. When a hypothetical patient is created, a future state is randomly chosen by a functional distribution with the best-fit parameters by the data set. A method to select the next occurring event was chosen through sample times for each possible event and the minimum chosen, namely following the logic according to which event was first to happen. Accordingly, the duration of this disease was also determined for this patient. An explicit simulation clock keeps track of the passage of time. If the future state is death, then this patient goes to the condition of the patient death box shown in Figure 4. Otherwise, the patient continues the course of the disease with age and a future state in the simulation is selected.

In all subsequent cycles, risk factors generate random effects through R_a and $H(t)$ such that a pair of health states (i, j) takes place at different times continuously for each patient. Moving forward with time, the simulation mimics disease progression by creating diabetic complications one by one until death occurs, or the simulation is terminated depending on whichever comes first. By this procedure, the simulation can model the disease progressive pathway of an individual. The micro-simulation was conducted using a joint programme written in Python and Excel. Accordingly, each hypothetical patient generates their trajectories over time by simulation. The simulation was conducted with 12,000 patients over the time horizon to produce cumulative incidences consisting of data corresponding to observations of NHIRD. While running the simulation, the model was verified by comparing the simulation outcomes with the observations. All simulation runs were used to reach convergence in outcomes, with an ordinary least-square model used to fit the T2DHoc predicted incidence rates to observed incidence rates, and slope, intercept, and R^2 were used to evaluate prediction accuracy. After the simulation experiments with the expected risk factors in representative populations generated hypothetically from the patient characteristics, a set of diabetes durations and a reasonable level of complications were collected and used to predict the diabetic risks of diabetic patients with the same conditions.

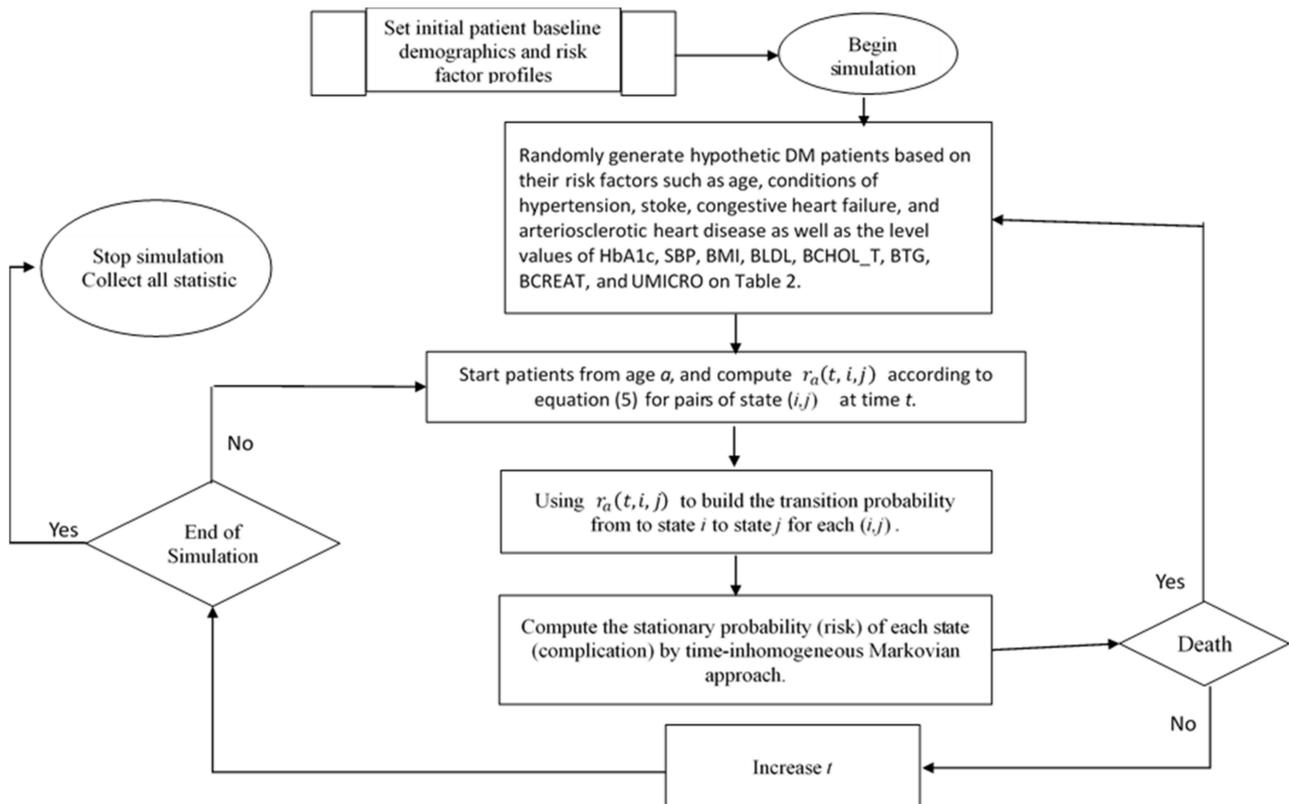


Figure 4. Flowchart of nonhomogeneous Markovian simulation for diabetes mellitus complications.

2.5. Internal and External Validation

According to the best practice recommendations by the International Society for Pharmacoeconomic Outcomes Research and the Society for Medical Decision-making Task Force [29,30], model transparency and validation, in particular, are most important. T2DHoc maintains transparency, since its development begins on the first day. Public health data on which the model is based are always available on request by anyone who has permission issued by NHIRD. In addition, the model is open to internal researchers who are investigators of NHID projects. The model validation involved face validity, verification, cross-validity, external validity, and predictive validity. For a general face validation of this model, we consulted with diabetologists/endocrinologists to ensure that the model was constructed to reflect and be used according to their expectations. Model verification was performed to ensure that there were no unintentional computational errors and each equation as well as the programme codes were checked by the model implementation.

In general, internal validation is designed to assess whether the model output is internally consistent with the studies of the data sources used to model the disease progression. The model construction with an interaction framework was examined by endocrinologists, nephrologists, cardiologists, and other related field experts. Moreover, the model equations with parameters were selected through sufficient statistical tests conducted by the SAS software. The model framework and assumptions were examined, and the definition, as well as the use of parameters embedded in each risk equation, were reviewed. After face validation and functional specification as described above, model verification was executed and evaluated by comparison using Equation (5), where the parameter calibration was applied for numerical stability. The calculation of cumulative incidence in the competing risk analysis was carried out by computation in Python.

3. Research Findings

3.1. Model and Patient Characteristics

The characteristics of diabetic patients are shown in Table 1. In summary, 163,452 T2D patients were included; 44.40% were female. The mean diagnosed age and diabetes history were 54.00 ± 11.86 (range: 18.51, 98.07) and 5.56 ± 6.28 (range: 0, 61.51) years, respectively. The mean baseline HbA1c level was $7.8 \pm 2.10\%$; body mass index ($26.481 \pm 3.97 \text{ kg/m}^2$), triglycerides ($172.64 \pm 135.51 \text{ mg/dL}$) and LDL ($116.13 \pm 35.28 \text{ mg/dL}$) were slightly higher than the normal range; systolic ($130.65 \pm 18.16 \text{ mmHg}$) and diastolic ($79.77 \pm 13.82 \text{ mmHg}$) blood pressures were within the normal ranges. Patients with a history of diabetic complications and risk factors were identified and collected. In the simulation, the factors and biomarkers were labelled as shown in Table 2. The patient disease pathways over ten years were collected as completely as possible to study the natural course of prolonged T2D complications and inconsistent fragmented data, and solely trace patients first enrolling P4P programme between 2002 and 2005 were excluded. In total, 12,242 patient pathways with at least one complete pathway were identified. The data collection procedure is outlined in Figure 5.

Table 2. General characteristics of T2D patients in NHIRD of Taiwan.

| Regression Model Coefficients | Mean | Standard Deviation | Definitions/Values |
|-------------------------------|-------|--------------------|--|
| AGE | 55.05 | 13.4 | Age in years at diagnosis of diabetes |
| Male | (1,0) | X | 1 for male; 0 for female |
| e_ht | (1,0) | X | 1 for history of Hypertension; 0 otherwise |
| e_cva | (1,0) | X | 1 for a history of Stroke; 0 otherwise |
| e_chf | (1,0) | X | 1 for a history of Congestive Heart Failure; 0 otherwise |
| e_ashd | (1,0) | X | 1 for a history of Arteriosclerotic Heart Disease; 0 otherwise |
| HBA1C | 7.84 | 1.99 | HbA1c (%), 5-year moving average of monthly values |
| Base_SBP | 126.1 | 31.1 | Systolic blood pressure at diabetes diagnosis (mm Hg) |
| BMI | 26.4 | 4.4 | Body mass index at diabetes diagnosis (m/kg^2) |
| BLDL | 116 | 36.9 | Low-density lipoprotein at diabetes diagnosis (mg/dL) |
| BHDL | 44.8 | 13.87 | High-density lipoprotein at diabetes diagnosis (mg/dL) |
| BCHOL_T | 191 | 53 | Total cholesterol at diabetes diagnosis (mg/dL) |
| BTG | 174 | 139 | Triglycerides at diabetes diagnosis (mg/dL) |
| BCREAT | 0.9 | 0.46 | Serum creatinine at diabetes diagnosis (mg/dL) |
| UMICRO | 50.1 | 377 | Urine microalbumin diabetes diagnosis (mg/dL). |

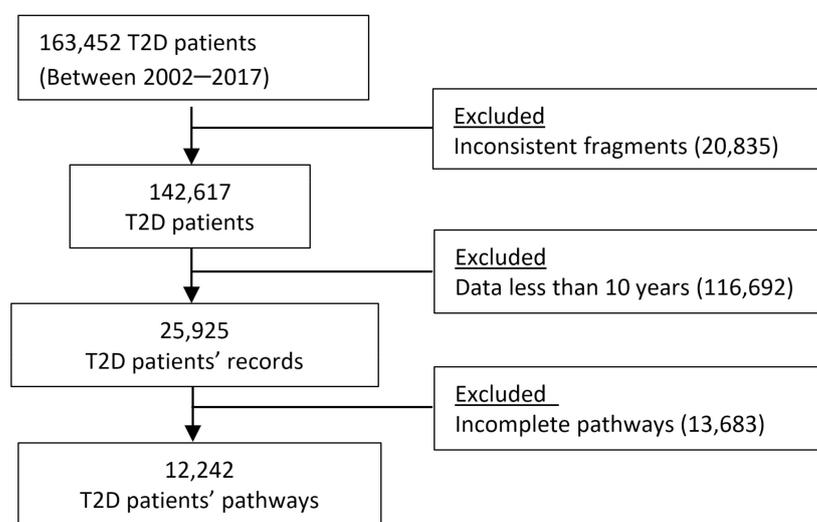


Figure 5. Process of identifying eligible patients' pathways in analysis. Abbreviation: T2D, type 2 diabetes mellitus.

3.2. Processed and Final Outcomes

The changes in HbA1c levels or at different time points can have different implications for the clinician or diabetic complications [19]. As one of the prediction functions in our investigation, a level function HbA1c of time was built depending on patient age (Figure 3). Among the possible diabetic risks, the total number of initial complications identified included 2015 cases with ASHD (16.5%), 827 cases with CHF (6.8%), 549 cases with ISC (4.5%), 2219 cases with FESRD (18.1%), 246 treatments for a detached retina or vitreous haemorrhage and 7241 all-causes, with recurrent events including 284 cases with ASHD (14.1%), 59 cases with CHF (7.1%), 22 cases with ISC (4.0%), 2924 cases with IHD, and 1158 cases with ESRD (Tables 3 and 4). However, if we consider the complication risks within 5 years, then the number of events occurring at the first layer and the second layer are shown in Tables 5 and 6.

Table 3. Incidence proportion of study complications in the patients with type 2 diabetes mellitus within 15 years.

| ASHD | FESRD | CHF | ISC | EYE | FIN_FOOT |
|-------|-------|------|------|------|----------|
| 16.5% | 18.1% | 6.8% | 4.5% | 2.0% | 1.0% |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

Table 4. The recurrence proportion of study complications in type 2 diabetes mellitus within 15 years.

| RE_ASHD | RE_CHF | RE_ISC |
|---------|--------|--------|
| 14.1% | 7.1% | 4.0% |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; ISC, ischemic stroke.

Table 5. Incidences of first layer complications or death in type 2 diabetes mellitus over 5 years.

| First Layer after DM | | | | | |
|---------------------------|----------------|-------------------|------------------|----------------------|------------------------|
| Complication Observations | ASHD 1363 | CHF 308 | ISC 415 | FESRD 2073 | EYE 218 |
| Complication Observations | FIN_FOOT 96 | ASHD + CHF 410 | ASHD + ISC 69 | CHF + ISC 20 | ASHD + CHF + ISC 14 |
| Complication Observations | FESRD 9 | CHF + FIN 3 | DEATH 7241 | Observations: 12,242 | |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; DM, diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

Table 6. Re-occurrences of study complications or death in type 2 diabetes mellitus over 5 years.

| Second Layer after DM | | | | | |
|---------------------------|----------------|---------------|---------------|--------------------|-----------|
| Complication Observations | ASHD 315 | CHF 100 | ISC 41 | FESRD 131 | EYE 46 |
| Complication Observations | FIN_FOOT 14 | FESRD 1036 | DEATH 1128 | Observations: 2821 | |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; DM, diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

3.3. Internal Validation

- (1) Taking ASHD as an example, the analysis showed that the proportion of diabetes complicated by ASHD at 55–60 years of age varies as age increases from 0.05 to 0.1. A comparison of the actual incidence of ASHD and the simulation ratio was almost the same at 55–56 years old; 57–58 is slightly lower but still within the prediction interval, and the actual incidence of complications at 59–60 years old is higher than the simulated ratio.

- (2) Comparison and display of the simulation analysis
 - (a) The first actual concurrent ASHD and simulation results were 2015 and 2268 patients, with an annual rate of 0.0114 and 0.0128, respectively, there were 284 relapses and 309 patients, with an annual rate of 0.0016 and 0.0017, respectively.
 - (b) The first actual concurrent stroke and simulation results were 549 and 594 patients, with an annual rate of 0.0031 and 0.0033, respectively, there were 22 relapses and 29 patients, with an annual rate of 0.0001 and 0.0002, respectively.
 - (c) The first actual concurrent CHF and simulation results were 828 and 780 patients, with an annual rate of 0.0047 and 0.0044, respectively.
 - (d) The first actual concurrent renal failure and simulation results were 2250 and 2268 patients, with an annual rate of 0.0127 and 0.0128, respectively.
- (3) A comparison of the simulated incidence rates of the first complication and overall deaths with observations in 10 years is presented in Figure 6. From the observational data, the prediction gap of FESRD was the smallest, e.g., underestimated 0.3%, and the prediction gap of retinopathy the largest, e.g., overestimated by 3.2%. Most of the predicted complication rates were higher than the actual rates, with gaps of 1.30% for death, 2.60% for ASHD, 3.20% for retinopathy, and 2.90% for ASHD + CHF, 2.00% difference in infarct stroke, 1.80% difference in CHF, 1.70% difference in diabetic amputation, 0.00% difference in CHF + infarct stroke, 0.10% difference in CHF + diabetic amputation. The difference between atherosclerosis heart disease + infarct stroke was 0.20%. A similar accuracy is shown when further applying it to predict recurrent complications (Figure 7).
- (4) The distribution of prediction error by complications and death rate is shown in Figures 8 and 9. The error ranges between the overall simulation results, and the actual observations were within 5% of the predicted value of complication and death rate and the observation quality. Suppose the average absolute percentage error was used for evaluation, and only 5 out of 10,000 events included. In that case, ASHD, death, and ESRD are within the generally accepted 30% range. Moderate complications are the incidence of stroke and CHF, with the incidence of foot lesions being highly overestimated.

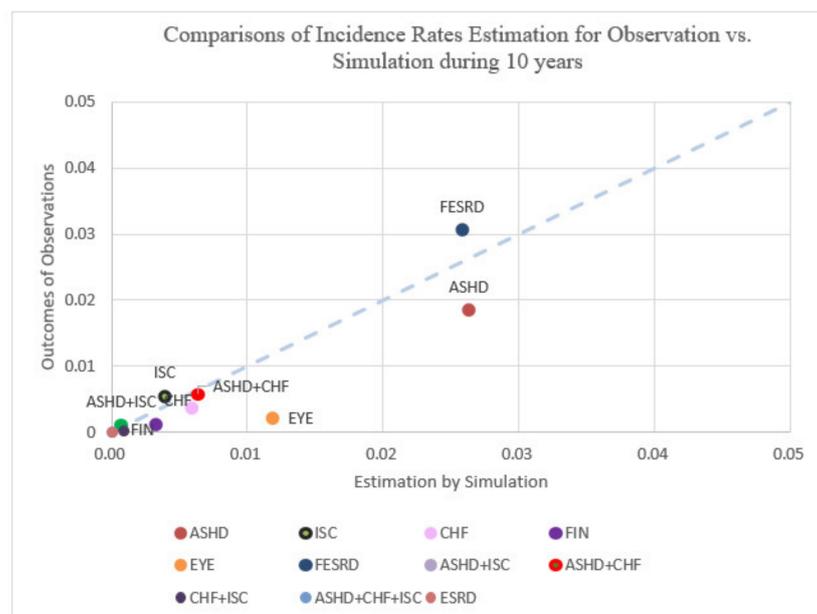


Figure 6. Comparisons of incidence rates estimation for observation vs. simulation over the course of 10 years.

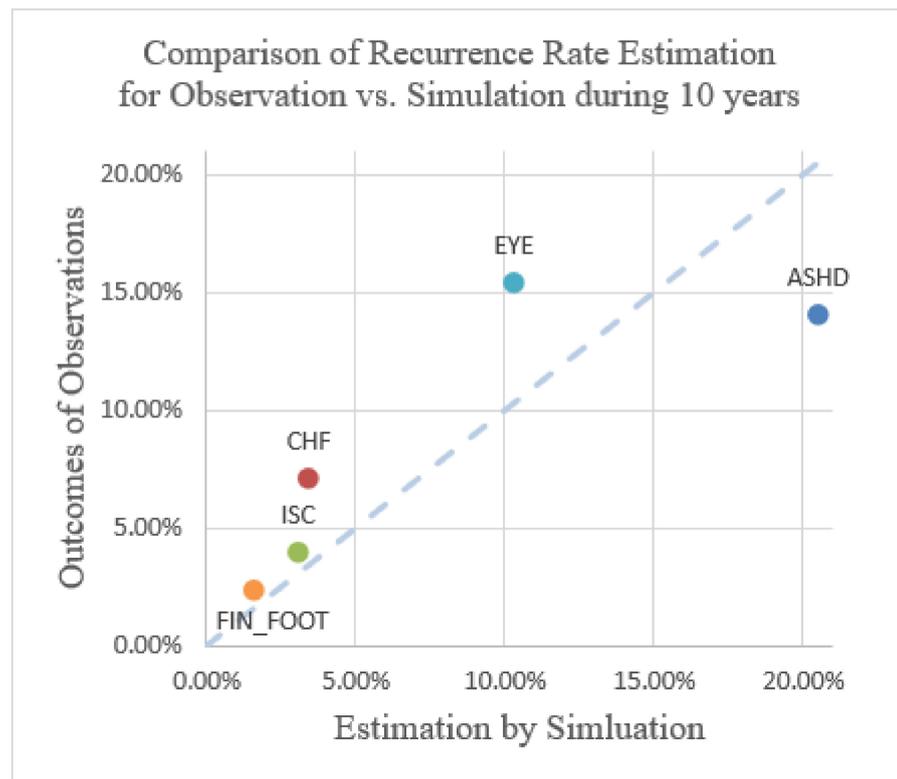


Figure 7. Comparison of recurrence rate estimation for observation vs. simulation over the course of 10 years.

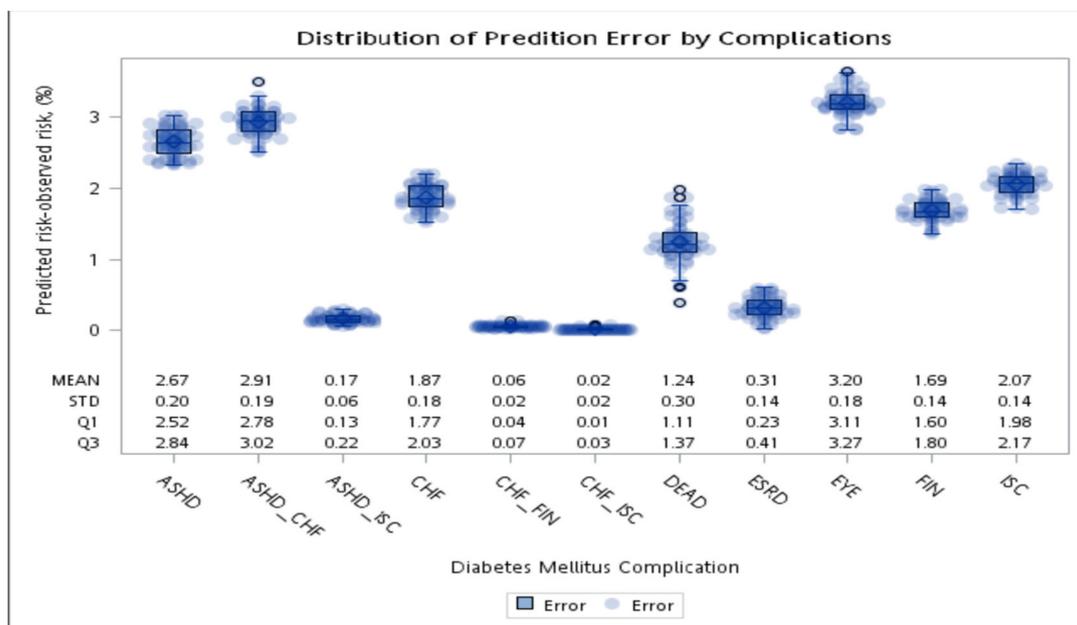


Figure 8. The average absolute error of the first event (complication or death) after diabetes diagnosis between the simulation and observations in 10 years.

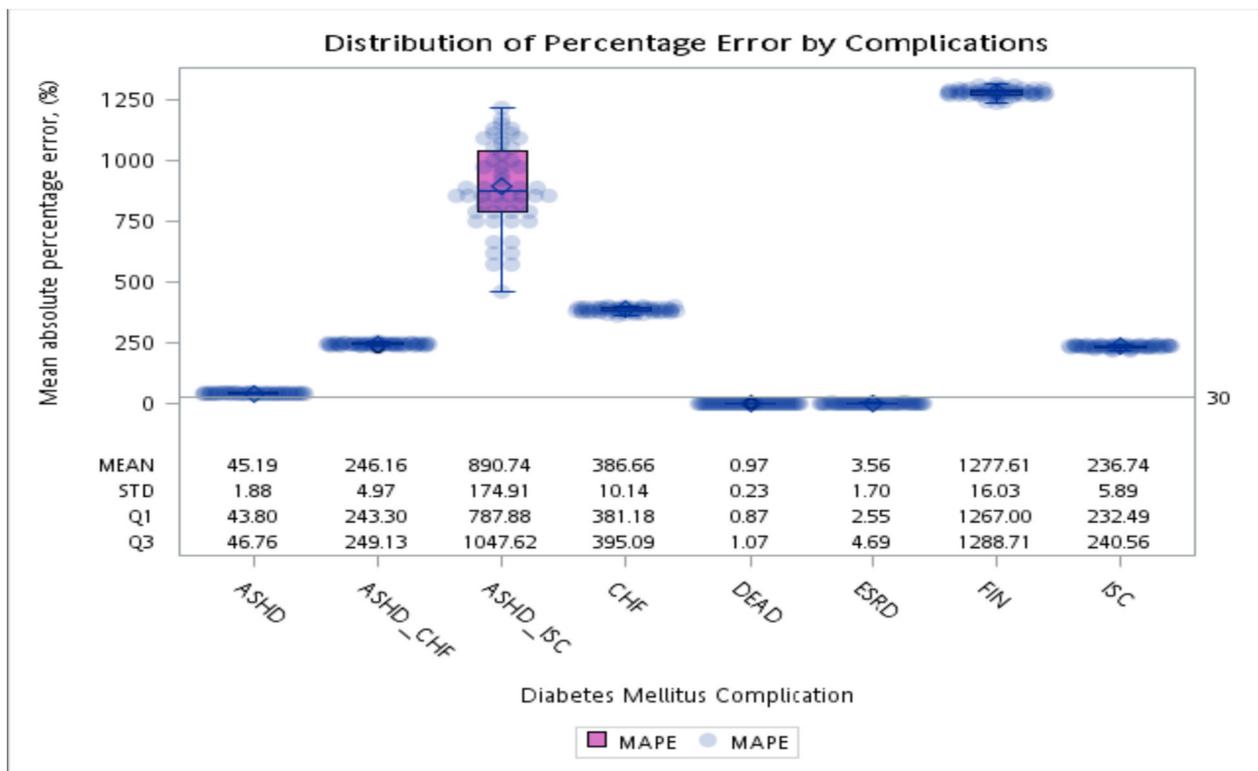


Figure 9. Average absolute percentage error of the first event (complication or death) after diabetes diagnosis between the simulation and observations in 10 years.

For example, the model estimates the cardiovascular risk from the T2DHoc risk engine, reflecting the disease progression of T2D patients in the P4P programme. The internal validation of the current T2DHoc risk engine can be shown by checking 5-year occurrences of each complication after the diagnosis of diabetes in Table 7. For further validation for a long-term prediction, we examine the incidence rate and recurrence of risks within 10 years. In a long-term prediction, ASHD+CHF with the event rate predicted by the risk engine was close to the observed rate, indicating that the model accurately predicts; the prediction confidence intervals may vary with different complications depending on health conditions, but the model has captured the disease progression trend.

Table 7. Simulation results versus observations for study complication in type 2 diabetes mellitus over 5 years.

| Complication | ASHD | CHF | ISC | FESRD | EYE | DM + Unclassified |
|--------------------------|----------|----------|-------|---|-----|-------------------|
| Estimation by Simulation | 1940 | 391 | 364 | 1199 | 646 | 60 |
| Observation | 1363 | 308 | 415 | 2073 | 218 | 118 |
| Complication | FIN_FOOT | ASHD+CHF | DEATH | | | |
| Estimation by Simulation | 109 | 491 | 7033 | 12,233 hypothetical patients in a 5-year simulation | | |
| Observation | 96 | 410 | 7241 | 12,242 patients in observation in 5 years | | |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; DM, diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

3.4. External Validation

External validation was conducted by comparing different models, such as UKPDS [14], Osaka model [31], the Korean model, and the Hong Kong model, by comparing the risk events predicted by the model, with observed clinical outcomes from research studies that were not directly used to inform disease progression.

- (1) The well-known Japanese diabetes literature [31] regarding T2D clinical trials from 1995 to 1996 involves a total of 2205 people aged 40–70 years old with HbA1c>6.5% who were randomly assigned to a lifestyle intervention group and conventional treatment group. All patients have initial data for both groups. Two sets of initial data were used in the Taiwan diabetes model to simulate 5000 hypothetical patients and compare the complication rates. From the literature, after 7.8 years of follow-up, the incidences of complications in the intervention group were coronary heart disease: 7%; stroke: 10%; nephropathy: 6.7%; eye lesions: 29% and in the control group was arterial heart disease: 7%; stroke: 6.5%; nephropathy: 6.7%; eye disease: 35.7%. Simulation results of the intervention group were coronary ASHD: 10.3%; ISC: 2.6%; ESRD: 1.1%; eye disease 3.6% and the simulation results of the control group were coronary ASHD: 10.9%; ISC: 2.0%; ESRD: 0.8%; eye disease: 3.8%. Since this generation of nephropathy was defined as proteinuria (UACR>300 mg/g), and ocular lesions were defined as the diagnosis of clinical tests (phases 1–4), it is different from the diagnosis code and the definition of dialysis in the Taiwan diabetes model, so the observed values were much higher than the predicted values.
- (2) South Korea collected 732 diabetic generations from Boramae Hospital in 2006 for 6 years [32]. It was observed that 43 (6.6%) patients developed coronary heart disease, and the use of the UKPDS risk formula would lead to the overestimation of the disease risk of patients. The patients' initial data were used in the Taiwan diabetes model to simulate 5000 hypothetical patients and estimate coronary ASHD for 6 years, and the result was 6.8%, consistent with the observed ratios.
- (3) Diabetes generation clinical trials in Hong Kong were developed in 1995, with a total of 7534 patients with T2D collected (the average course of diabetes was 7.1 years, and the prevalence rate of hypertension was 70%) [33]. They were tracked for 5 years and have targeted different common types of diabetes. According to the literature, the numbers of major complications in this generation within 5 years were death: 763; coronary heart disease: 377; stroke: 362; diabetic nephropathy: 693; CVD: 1120; ESRD: 282 and the percentage of major complications was death: 10.13%; coronary heart disease: 5%; stroke: 4.8%; diabetic nephropathy: 9.2%; CVD: 14.87%; end-stage renal disease: 3.74%. These data were used in the Taiwan diabetes model for simulation revealing death: 1862; coronary ASHD: 1142; ISC: 446; ESRD: 445, whereas the percentage of major complications was death: 17.2%; coronary ASHD: 11.4%; ISC: 4.46%; ESRD: 4.5%. The death rate by simulation was slightly higher than the observed values. The predicted rate of stroke, coronary ASHD, and ESRD was slightly higher than the observed rates.

The external validation studies selected in this study represent a broad range, including observations of each complication risk with duration, as well as collected data from disease progression. The specific external validation studies included in this analysis were: National Health dataset, clinical datasets. The external validation has shown the comparison of the T2DHoc and UKPDS by 3867 UK patients which is illustrated in Figure 10. The UKPDS results mentioned above also support the use of the current risk engine.

Data from the literature were compiled for comparing the discrete event simulation for advanced Asian countries, showing that subjects in South Korea and Japan were older than subjects in Taiwan. The mean values of HbA1c, SBP, and HDL cholesterol were lowest in South Korea and highest in Japan. The mean value of triglycerides was lowest in Japan and highest in Taiwan. Limited information on LDL cholesterol in Japan and South Korea and limited information on serum creatinine and urine albumin in Japan, South Korea, and Hong Kong were noted. The simulation data of the three countries are hypothetically assumed to be the same as the Taiwanese values. We further compared the main complications listed in this plan to the original definition of diabetes in Asian countries. The definition of complications for Japanese diabetics mainly uses clinical diagnostic criteria to define major complications, while Hong Kong defines complications, compared to Taiwan, are based on diagnosis codes. South Korea's definition of complications is

based on retrospective medical history. Based on the differences in the four definitions, the Taiwan model was found to be more comparable in terms of the definition of ASHD for the diabetes generation in other advanced Asian countries. Hong Kong is better than Japan in the sense of advanced research on diabetic complication development, and Japan is better than South Korea. The comparability of the definition of stroke and nephropathy is that Hong Kong is better than Japan. It is worth noting that the ratio of death to ASHD model predicted value is much higher than the observed value, which may be related to medical ecology, and the actual cause remains to be clarified.

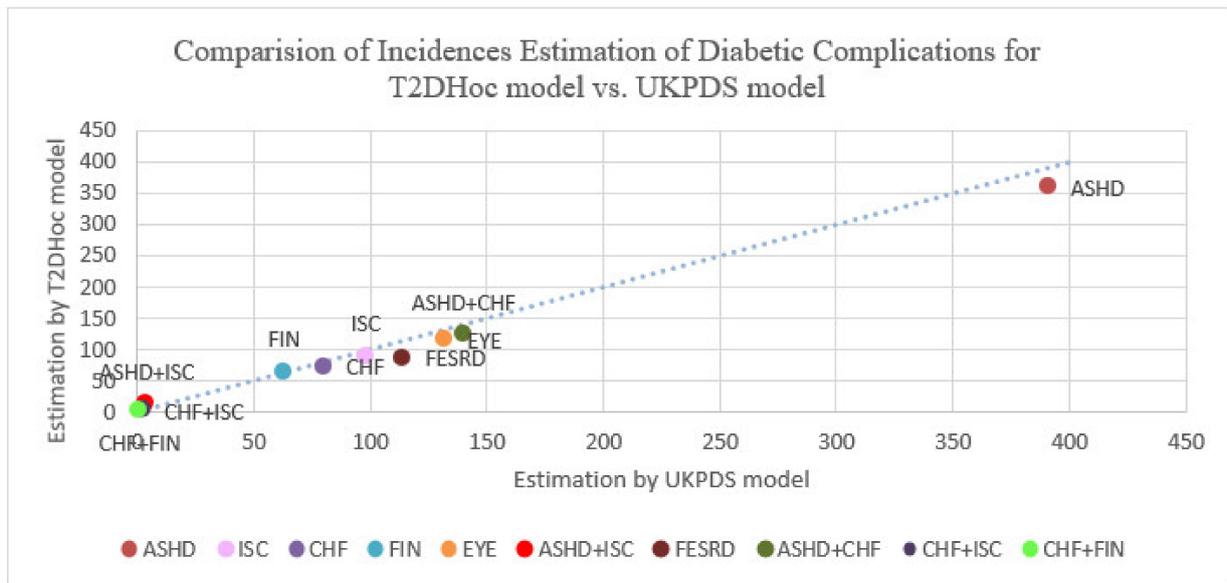


Figure 10. Comparison of incidence rates estimation of diabetic complications for UKPDS vs. T2DHoc models with 3867 hypothetical patients in 10 years. Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; T2D, type 2 diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

4. Discussion

The study successfully extracted information from a long-lasting trial, with more than 163 thousands participants following up 16 years. A good predicted complication and mortality rates were approved against the setting observed numbers of complications or death in the original cohort. In addition, better and optimal predictabilities were further confirmed when compared to the UKPDS Outcomes model and applied the model to other Asian populations, respectively.

The T2DHoc is based on more comprehensive follow-up data as it captures more outcomes has significant advantages over existing models, and more comprehensively captures the progression of diabetes. It permits detailed and reliable lifetime simulations of key health outcomes in people with T2D, especially for Asian societies. The predicted event rates for complications by the T2DHoc model are slightly lower than most existing models, despite estimated rates in those models often being overshoot. The results suggest that applying recently developed models to clinical practice should pay some attention to their intrinsic variations such as data source, operational definitions, time-span, complexity, and feasibility. Undoubtedly, our model derived from practical data shows a large promise to shrinkage the abovementioned gaps. However, more research on methods in selecting models by clinically usefulness index is needed.

Previous studies have pointed out that T2D models mainly developed using Western patients do not appropriately reflect the risks of CVD and renal disease in the Asian population [34,35]. It is reasonable to suspect that ethnicity, genes (ALDH2 deficiency), environmental factors (birth cohort effects: World Wars, PM 2.5 pollutants, food habits),

as well as healthcare systems are highly involved in the accuracy of a prediction model [36]. Although the definitions of selected complications are different, the T2DHoc model shows better performances than the UKPDS Outcomes model or Framingham risk score when applying them to the Asian T2D population. Further studies to compare performances of current models developed in the Western population and ours in more Asian ethnicities such as Australian T2D patients are required.

The T2DHoc model has several advantages. First, it simultaneously takes many major complications, absorbed status (death), and their dynamic time relationship within a single analytic framework. Thus, information on complications is collected more comprehensively. Then, the dynamic relationships between age and HbA1c levels were established and incorporated into the model structure, allowing more flexibility for practical applications. Finally, the follow-up time in the current study is obviously longer than for other models developed from data collecting by clinical trials. The accuracy for complications that need more time to be observed, such as renal failure, should be largely improved.

Some limitations of the study should be recognized. The study mainly takes laboratory data at baseline rather than consequent medication change for the model development, which makes us not able to understand the influences of medication use on the probabilities of health state change. For example, patients who used cardiovascular prevention drugs (such as antiplatelet drugs, renin-angiotensin-aldosterone blockade, and beta-blocker) may change the probability of experiencing the first ASHD (health state j) from baseline (health state i). Although adding information on medication usage may enhance the current model's performance, medication use changes over time and is modified depending on the patient's conditions, making our model more difficult to understand and practically apply. In addition, due to a long observed time, we could not rule out and quantify the influences of the T2D care improvements (new drugs) over time on our results.

5. Conclusions

Model building is an iterative process, and models need to be updated as new information becomes available. The T2DHoc was based on P4P data collected from 2002 up to now. Additional information collected during the P4P 10-year period provided an opportunity to update the simulation model and to incorporate data on new risk factors and outcomes that were unavailable when other diabetes prediction models were constructed.

The progression of diabetes in Asian patients is different from that reported in western countries. For Asian patients, ASHD, chronic kidney disease, and stroke are the most likely first complications to occur than CHF, ASHD, retinopathy, and amputation. Risks of developing a further complication vary according to patients' existing complication profiles. Patients with an existing cardiovascular complication or retinopathy have a higher risk of developing ASHD and chronic kidney disease. These results inform clinical decision-making regarding prioritising monitoring and interventions for diabetic patients who are at high risk of developing severe complications in this region.

The study demonstrates how a simplified whole disease model approach can be applied to rare diseases to provide the disease burden estimation, serving as an information tool for payers, for diseases with small patient volume and an unknown cost burden. Additionally, machine learning, data mining, and text mining can be applied to data contained in electronic health records to further research other diseases.

Though efficacy and safety were not taken into account in this model, we believe that it still provides a good understanding of resource utilisation and costs of conditions with poorly documented epidemiology and disease burden. The enormous economic burden of type 2 diabetes mellitus T2D can be reduced by implementing inexpensive, easy-to-use interventions, such as joining a diabetes prevention programme.

T2DHoc is a prediction model with a single analytic framework suitable for predicting T2D risks, both for long-term and short-term complications. This pilot study demonstrated that non-homogeneous-based modelling is useful for T2DHoc modelling. Further research,

including the most updated treatments, should be sustained to complete the disease process of T2D with diabetes management.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/info12080326/s1>, Table S1. Probabilities of diabetic complications in each cycle after computation.

Author Contributions: Conceptualization, M.-Y.L., Y.-W.C. and S.-J.H.; methodology, M.-Y.L., P.-H.W. and H.L.; validation, M.-Y.L., P.-H.W., Y.-W.C. and S.-J.H.; formal analysis, J.-S.L. and T.-Y.H.; resources, M.-Y.L., C.-C.H. and H.L.; data curation, J.-S.L., P.-H.W. and T.-Y.H.; writing—original draft preparation, M.-Y.L.; writing—review and editing, P.-H.W., C.-C.H. and H.L.; visualization, M.-Y.L., T.-Y.H. and Y.K.; supervision, S.-J.H., Y.K., C.-C.H. and H.L.; project administration, M.-Y.L. and H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partly supported by the Taiwan Ministry of Science and Technology (most: 105-2314-b-037-065-, 106-2314-b-002-253-, 107-2314-b-037-122- and 108-2314-b-037-110-), Kaohsiung Medical University Hospital (kmuh: 105-5r17 and 106-6r21), and Center for Big Data Research, Taiwan National Health Research Institutes (nhri-ex107 -10717pi and nhri-ex108-10505pi).

Institutional Review Board Statement: This study was approved by the institutional review boards of Kaohsiung Medical University Hospital (kmuhirb-e(i)-20160070), National Chengchi University (nccu-rec-201603-i005) and National Health Research Institutes (ec1050505-e). All study procedures were conducted by following the declaration of the Helsinki principle.

Informed Consent Statement: The ethical review boards waived patient consent because all data used in the study have been de-identified.

Data Availability Statement: The data used in our study were limited to research purposes only and cannot be made publicly available under the regulation of the personal information protection act in Taiwan. The raw data were obtained from the national health insurance administration, ministry of health and welfare https://www.nhi.gov.tw/content_list.aspx?N=2d2faf5214807829&topn=787128dad5f71b1a (accessed on 21 May 2021) and can be made available to qualified researchers upon request.

Acknowledgments: We would like to thank the team members, Ming-Huan Chan, Shyi-Jang Shin, Jer-Chia Tsai, Kun Der Lin, Yi-Ting Lin, Meng-Chuan Huang, Hui-Min Hsieh, Yi-Hsin Yang, Li-Jen Cheng, Sheng-Tai Huang, Teng-Hui Huang, and Shu-An Yang for providing their valuable opinions and devoting their efforts on our model development.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Clinical Definition of Diabetic Complications

- Retinopathy
 - Diabetic retinopathy is a primary cause of blindness worldwide, and this serious complication of diabetes is already present at the time of clinical diagnosis of type 2 diabetes in some patients.
 - In the Wisconsin Epidemiologic Study of Diabetic Retinopathy, 3.6% of patients with type 1 diabetes and 1.6% of patients with type 2 diabetes were blind.
 - It is recommended that patients with type 2 diabetes have an initial comprehensive eye examination by an ophthalmologist or optometrist shortly after being diagnosed with diabetes.
- Neuropathy
 - Diabetic peripheral neuropathy is frequent, and 50% of people with type 2 diabetes have neuropathy and therefore are at risk of developing diabetic foot ulcers.
 - Diabetic neuropathy is known by the American Diabetes Association as “the presence of symptoms and/or signs of peripheral nerve dysfunction in people with diabetes after the exclusion of other causes.” A foot ulcer is one of

the major complications in patients with diabetes, with a 15% lifetime risk of amputation.

- Nephropathy
 - Diabetic nephropathy is the leading cause of renal failure in the United States.
 - The kidneys begin to leak, and albumin passes into the urine. This can be preceded by lower degrees of proteinuria or microalbuminuria and can proceed to renal failure in the worst case.
 - Identification of people at high risk of rapid decline in renal function is important, and evidence-based interventions have been shown to prevent or slow the development toward advanced stages of nephropathy.
- Heart Disease
 - Diabetes is a well-known risk factor for coronary heart disease. Diabetes adds an about 2-fold risk for a wide range of vascular diseases, independently of other conventional risk factors.
 - Much research has been conducted to develop predictive models or risk scores for at-risk individuals from the general population. One of the best models is the Framingham score ([link](#)), which has been widely accepted and includes diabetes as a predictor.
- Hypoglycemia
 - People with type 1 diabetes often experience episodes of hypoglycaemia because they need to reduce the level of blood sugar by using insulin. Additionally, patients with type 2 diabetes may experience episodes of hypoglycaemia because of the increasing use of insulin in this group.
 - The fear induced by hypoglycaemia is pronounced, and the clinical results of this condition are serious. The literature suggests that the incidence of hypoglycaemia requiring emergency assistance reaches 7.1% per year among patients with diabetes and that as many as 6% of all deaths in patients with type 1 diabetes are due to hypoglycaemia.
- Insulin-Associated Weight Gain
 - In most patients with type 2 diabetes, it will eventually be necessary to begin insulin treatment to achieve the therapeutic goal of HbA1c < 7 mmol/L (126 mg/dL). The problem of weight gain induced by insulin has long been documented as an issue in diabetes treatment.
 - In the Diabetes Control and Complications Trial (DCCT), the average weight gain of patients with type 1 diabetes undergoing intensive treatment was 5.1 kg compared with 2.4 kg in the standard treatment arm, and similar results are seen for type 2 diabetes.

Appendix B. An Explanation of Markovian Approach

Appendix B.1. The Basic Computational Model

In this appendix, the Markov modelling approach is applied to consider the progression of diabetes disease with complications. Here, the primary purpose is to introduce the mathematical notations and formulae used in the paper. For detailed treatments to these topics, they may be referred to [27].

Consider a sequence of random variables $Y = \{Y_n, n \in N \cup \{0\}\}$ defined on a probability space (Ω, \mathbb{F}, P) with a finite set $\{s_0\} \cup E = \{s_1, s_2, \dots, s_m\}$ for $m < \infty$, where N is the set of all positive integers. In our case, a unit cycle is defined as one month and $E = \{\text{ASHD}, \text{CHF}, \text{ISC}, \text{FESRD}, \text{EYE}, \text{FINFOOT}, \text{ASHD} + \text{CHF}, \text{ASHD} + \text{ISC}, \text{CHF} + \text{ISC}, \text{ASHD} + \text{CHF} + \text{ISC}, \text{ESRD}, \text{DEAD}\}$. The state symbols are defined and shown in Table A1.

Table A1. Symbols of the states.

| Disease | State Symbol |
|------------------|--------------|
| DM | s_0 |
| ASHD | s_1 |
| CHF | s_2 |
| ISC | s_3 |
| FESRD | s_4 |
| EYE | s_5 |
| FIN_FOOT | s_6 |
| ASHD + CHF | s_7 |
| ASHD + ISC | s_8 |
| CHF + ISC | s_9 |
| ASHD + CHF + ISC | s_{10} |
| ESRD | s_{11} |
| CHF + FIN_FOOT | s_{12} |
| DEATH | s_{13} |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; DEAD, death; DM, diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

According to Grossetti et al. [27] and Equations (1)–(4) in our paper, a risk equation from state i at age a to state j during time t is defined in Equation (5). For example, a risk equation from ASHD to j , $j \in E$, where $t \in (t_0, t_1)$

$$r_a(t, ASHD, j) = \left(1 - \exp\left(\sigma(ASHD, j) \times \left((t - 1)^{c(ASHD, j)} - t^{c(ASHD, j)}\right) \times C_a(t, ASHD, j)\right)\right)$$

where $\sigma(ASHD, j)$ denotes the scale and $c(ASHD, j)$ denotes the shape of Weibull distribution from ASHD to j , respectively. Table A2 illustrates those parameters of Weibull distributions that have been excerpted from NHIRD associated with all possible states for the T2DHoc model. In addition, $C_a(t, ASHD, j)$ is computed by $C_a(t, ASHD, j) = \exp(R_a(t, ASHD, j))$, where

$$R_a(t, ASHD, j) = \sum \beta_i(j)x_i + \beta_{HbA1c}(j) \times x_{HbA1c}(a, t) + \beta_{age}(a, t, j).$$

Table A2. Weibull distribution with associated parameters from diabetes mellitus to different states.

| DM--> | Distribution | Parameter | Symbol | Estimate |
|------------------|--------------|-----------|--------|----------|
| ASHD + CHF + ISC | Weibull | Threshold | Theta | 416.999 |
| ASHD + CHF + ISC | Weibull | Scale | Sigma | 253.4067 |
| ASHD + CHF + ISC | Weibull | Shape | C | 0.339735 |
| ASHD + CHF + ISC | Weibull | Mean | | 1833.863 |
| ASHD + CHF + ISC | Weibull | Std Dev | | 5953.267 |
| ASHD | Weibull | Threshold | Theta | -24.101 |
| ASHD | Weibull | Scale | Sigma | 842.6903 |
| ASHD | Weibull | Shape | C | 1.664991 |
| ASHD | Weibull | Mean | | 728.9132 |
| ASHD | Weibull | Std Dev | | 464.655 |
| ASHD + CHF | Weibull | Threshold | Theta | -23.4148 |
| ASHD + CHF | Weibull | Scale | Sigma | 864.1454 |
| ASHD + CHF | Weibull | Shape | C | 1.669587 |
| ASHD + CHF | Weibull | Mean | | 748.6106 |

Table A2. Cont.

| DM--> | Distribution | Parameter | Symbol | Estimate |
|----------------|--------------|-----------|--------|-----------|
| ASHD + CHF | Weibull | Std Dev | | 475.191 |
| ASHD + ISC | Weibull | Threshold | Theta | -144.766 |
| ASHD + ISC | Weibull | Scale | Sigma | 1017.946 |
| ASHD + ISC | Weibull | Shape | C | 2.201152 |
| ASHD + ISC | Weibull | Mean | | 756.753 |
| ASHD + ISC | Weibull | Std Dev | | 432.3723 |
| CHF | Weibull | Threshold | Theta | 5.251415 |
| CHF | Weibull | Scale | Sigma | 743.1523 |
| CHF | Weibull | Shape | C | 1.547657 |
| CHF | Weibull | Mean | | 673.7453 |
| CHF | Weibull | Std Dev | | 441.0139 |
| CHF + FIN_FOOT | Weibull | Threshold | Theta | 125.999 |
| CHF + FIN_FOOT | Weibull | Scale | Sigma | 22.69363 |
| CHF + FIN_FOOT | Weibull | Shape | C | 0.178777 |
| CHF + FIN_FOOT | Weibull | Mean | | 7858.026 |
| CHF + FIN_FOOT | Weibull | Std Dev | | 180,164.4 |
| CHF + ISC | Weibull | Threshold | Theta | -2305.04 |
| CHF + ISC | Weibull | Scale | Sigma | 3287.634 |
| CHF + ISC | Weibull | Shape | C | 15.92766 |
| CHF + ISC | Weibull | Mean | | 875.576 |
| CHF + ISC | Weibull | Std Dev | | 245.5491 |
| DEATH | Weibull | Threshold | Theta | -43.0451 |
| DEATH | Weibull | Scale | Sigma | 921.1229 |
| DEATH | Weibull | Shape | C | 1.970673 |
| DEATH | Weibull | Mean | | 773.5216 |
| DEATH | Weibull | Std Dev | | 432.5532 |
| ESRD | Weibull | Threshold | Theta | 3.999 |
| ESRD | Weibull | Scale | Sigma | 56.41532 |
| ESRD | Weibull | Shape | C | 0.363986 |
| ESRD | Weibull | Mean | | 252.7458 |
| ESRD | Weibull | Std Dev | | 919.6434 |
| EYE | Weibull | Threshold | Theta | 1.989741 |
| EYE | Weibull | Scale | Sigma | 606.3383 |
| EYE | Weibull | Shape | C | 1.358656 |
| EYE | Weibull | Mean | | 557.3697 |
| EYE | Weibull | Std Dev | | 413.3502 |
| FESRD | Weibull | Threshold | Theta | -197.796 |
| FESRD | Weibull | Scale | Sigma | 1121.817 |
| FESRD | Weibull | Shape | C | 2.548924 |
| FESRD | Weibull | Mean | | 798.0508 |
| FESRD | Weibull | Std Dev | | 418.8006 |

Table A2. *Cont.*

| DM--> | Distribution | Parameter | Symbol | Estimate |
|----------|--------------|-----------|--------|----------|
| FIN_FOOT | Weibull | Threshold | Theta | -55.233 |
| FIN_FOOT | Weibull | Scale | Sigma | 909.3811 |
| FIN_FOOT | Weibull | Shape | C | 1.899072 |
| FIN_FOOT | Weibull | Mean | | 751.7311 |
| FIN_FOOT | Weibull | Std Dev | | 441.9696 |
| ISC | Weibull | Threshold | Theta | -15.9275 |
| ISC | Weibull | Scale | Sigma | 839.8184 |
| ISC | Weibull | Shape | C | 1.527653 |
| ISC | Weibull | Mean | | 740.5984 |
| ISC | Weibull | Std Dev | | 505.0916 |

Abbreviation: ASHD, arteriosclerotic heart disease; CHF, chronic heart failure; DM, diabetes mellitus; EYE, retinopathy; FESRD, first-time renal failure; FIN_FOOT, amputation; ISC, ischemic stroke.

Note that $\sum \beta_i(j)x_i$ is calculated with β_i and x_i , where β_i is chosen from the corresponding values in Table A1 and x_i is chosen from the corresponding value in Table 2. Similarly, $\sum \beta_{HbA1c}(j) \times x_{HbA1c}(a, t) + \beta_{age}(a, t, j)$ is computed with the associated glycated haemoglobin (HbA1c) function, which is defined by

$$x_{HbA1c}(a, t) = \begin{cases} \theta(a, t) + \varphi(a, t)x_{HbA1c}(a, t - 1) + \psi(a, t) \times t, & t < 60 \\ x_{HbA1c}(a, t - 1), & t \geq 60 \end{cases}$$

where show in Table A2.

Since the unit step transition probability matrix changes with age a , the proposed model clearly generates a nonhomogeneous Markovian process. With a given starting age and an initial healthy state, the probability associated with any likely realization consisting of chronically experienced states can be computed. Now, define the transition function $p_{i,j}^{t,s} = P(Y_s = j | Y_t = i)$ as a transition probability matrix with the elements $\mathbf{P}^{t,s} = (p_{i,j}^{t,s})_{i,j \in E}$. Namely, $P_{i,j}^{t,s}$ is the probability that a patient with the initial healthy state i at age written in terms of cycle time t will be at the health state j at age in terms of cycle time s . Specifically, when $s = t + 1$, $p_{i,j}^{t,t+1}$ in this matrix is the one-step transition probability from state i to state j with a starting time t and an ending time $t + 1$. That is for $i \neq j$ we have $p_{i,j}^{t,t+1} = r_t(t + 1, i, j)$ and $p_{i,i}^{t,t+1} = 1 - \sum_{k \neq j} p_{i,k}^{t,t+1}$ for $i = j$.

Let $\mathbf{P}(k)$ denote the unit step transition matrix with starting cycle time k and ending cycle time $k + 1$. Then, the transition probability matrix $\mathbf{P}^{t,s}$ can be computed by

$$\mathbf{P}^{t,s} = \prod_{k=t}^{s-1} \mathbf{P}(k)$$

where,

$$\mathbf{P}(k) = (p_{i,j}^{k,k+1})_{i,j \in E}$$

Table A3. The coefficients of linear regressions associated with each $\sum \beta_i(j)x_i + \beta_{HBA1C(j)} \times x_{HBA1C}(a, t)$.

| To | DEAD | | FESRD | | ASHD | | ISC | | CHF | | EYE | | FIN_FOOT | | ASHD + CHF | | ASHD + ISC | | CHF + ISC | | ASHD + CHF + ISC | | ESRD | |
|----------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|----------|----------|----------|--------|------------|--------|------------|--------|-----------|--------|------------------|--------|----------|----------|
| | β | st dev | β | st dev | β | st dev | β | st dev | β | st dev | β | st dev | β | st dev | β | st dev |
| age20 | -0.7789 | 0.1475 | -0.0260 | 0.1812 | -1.0272 | 0.4221 | -1.2432 | 1.0259 | -0.9531 | 1.0313 | -1.3342 | 1.0446 | 0.3192 | 0.6812 | -1.7127 | 1.1356 | -16.0922 | 1485 | -17.2909 | 6587 | -55.3862 | 5554 | 1.2294 | 1449.048 |
| age65 | 0.9191 | 0.0615 | 0.1533 | 0.1170 | 0.0835 | 0.1673 | 0.6429 | 0.2815 | 0.5981 | 0.3349 | -1.7501 | 0.7359 | -0.1431 | 0.5337 | 0.9108 | 0.2478 | -0.4589 | 0.8408 | 2.7913 | 1.3747 | 1.8459 | 1.6850 | -1.2748 | 1186.030 |
| age75 | 1.7633 | 0.0674 | 0.4914 | 0.1503 | 0.4381 | 0.2021 | 0.9106 | 0.3559 | 1.0436 | 0.3934 | -14.9763 | 847.6930 | 0.9108 | 0.5319 | 0.8869 | 0.3132 | 0.3527 | 0.7851 | -18.7117 | 33.476 | -21.0000 | 82.119 | -7.7770 | 2655.494 |
| male | 0.5092 | 0.0526 | 0.2445 | 0.0956 | 0.9019 | 0.1521 | 0.0768 | 0.2461 | -0.0822 | 0.2897 | -0.3124 | 0.3500 | 0.6981 | 0.4176 | 0.6349 | 0.2272 | 0.5823 | 0.6758 | 3.3709 | 1.7556 | 20.2201 | 5289 | -12.7064 | 2551.357 |
| e_ht | 0.0277 | 0.0583 | 0.7992 | 0.1141 | 0.6795 | 0.1696 | 1.0867 | 0.3641 | 0.6678 | 0.3901 | 1.6708 | 0.4581 | 0.5130 | 0.4558 | 0.2664 | 0.2770 | -1.1441 | 0.6901 | 21.6020 | 3265 | -3.9781 | 2.6120 | 13.8813 | 1470.210 |
| e_cva | 0.4707 | 0.0636 | 0.3209 | 0.1309 | 0.0826 | 0.1878 | 1.7122 | 0.2547 | -0.2178 | 0.4201 | 0.0224 | 0.6075 | 0.8527 | 0.4647 | 0.0514 | 0.2735 | 2.8948 | 0.6656 | 0.7688 | 1.2840 | 4.0271 | 1.8501 | -9.3843 | 1660.328 |
| e_chf | 0.5454 | 0.0786 | 1.1513 | 0.1303 | 0.5479 | 0.1939 | -0.0854 | 0.4423 | 2.2933 | 0.3171 | -0.6067 | 1.0243 | 0.6316 | 0.5776 | 2.0512 | 0.2371 | 0.4045 | 0.8412 | 1.4417 | 1.3454 | 4.7762 | 2.2172 | 5.2080 | 1459.719 |
| e_ashd | 0.0489 | 0.0607 | 0.1033 | 0.1116 | 1.5602 | 0.1371 | -0.0581 | 0.2880 | 0.1161 | 0.3162 | -0.1384 | 0.4878 | 0.6804 | 0.4197 | 1.1186 | 0.2255 | 2.1065 | 0.6522 | -0.6428 | 1.2976 | 1.7371 | 1.5976 | -18.1875 | 1087.827 |
| hba1c | 0.0436 | 0.0113 | 0.1420 | 0.0183 | -0.0023 | 0.0312 | 0.1163 | 0.0529 | 0.1019 | 0.0565 | 0.0851 | 0.0727 | 0.0734 | 0.0832 | 0.0893 | 0.0434 | -0.0903 | 0.1632 | 0.3321 | 0.2284 | 0.6072 | 0.2443 | -7.8867 | 178.588 |
| BASE_SBP | 0.0021 | 0.0013 | 0.0068 | 0.0009 | -0.0012 | 0.0036 | 0.0051 | 0.0053 | 0.0043 | 0.0065 | 0.0080 | 0.0047 | 0.0072 | 0.0024 | 0.0088 | 0.0026 | 0.0135 | 0.0162 | 0.0230 | 0.0281 | 0.0632 | 0.0381 | -0.2593 | 13.366 |
| bmi | -0.0569 | 0.0072 | -0.0534 | 0.0126 | -0.0288 | 0.0182 | -0.0310 | 0.0334 | -0.0653 | 0.0382 | -0.0574 | 0.0443 | 0.0027 | 0.0482 | -0.0921 | 0.0310 | -0.0234 | 0.0814 | 0.3129 | 0.0933 | 0.0970 | 0.1213 | -1.9589 | 97.178 |
| BLDL | -0.0056 | 0.0010 | -0.0007 | 0.0018 | 0.0072 | 0.0024 | 0.0045 | 0.0050 | 0.0051 | 0.0054 | -0.0019 | 0.0061 | -0.0044 | 0.0078 | 0.0022 | 0.0047 | -0.0072 | 0.0104 | -0.0194 | 0.0286 | 0.0026 | 0.0223 | -0.1735 | 9480 |
| BHDL | -0.0036 | 0.0021 | -0.0028 | 0.0039 | -0.0120 | 0.0059 | 0.0060 | 0.0104 | 0.0083 | 0.0118 | -0.0107 | 0.0135 | -0.0096 | 0.0159 | 0.0088 | 0.0091 | -0.0022 | 0.0237 | 0.0648 | 0.0351 | 0.0302 | 0.0558 | 0.6522 | 14,198 |
| BCHOL_T | 0.0019 | 0.0009 | 0.0021 | 0.0015 | -0.0013 | 0.0019 | -0.0016 | 0.0043 | -0.0041 | 0.0043 | 0.0050 | 0.0050 | 0.0022 | 0.0068 | 0.0015 | 0.0042 | -0.0058 | 0.0087 | 0.0049 | 0.0220 | 0.0334 | 0.0172 | -0.1649 | 6465 |
| BTG | 0.0000 | 0.0002 | 0.0007 | 0.0004 | 0.0010 | 0.0005 | 0.0011 | 0.0004 | 0.0011 | 0.0014 | -0.0007 | 0.0015 | -0.0028 | 0.0023 | 0.0007 | 0.0010 | -0.0015 | 0.0022 | -0.0061 | 0.0104 | -0.0114 | 0.0104 | 0.0367 | 1098 |
| BCREAT | 0.0915 | 0.0295 | 0.2103 | 0.0190 | 0.0628 | 0.0798 | 0.1896 | 0.0642 | 0.0350 | 0.2336 | -0.0261 | 0.4008 | 0.0259 | 0.2990 | 0.0091 | 0.1814 | -0.4221 | 0.9489 | -1.2522 | 2.0243 | 0.0100 | 1.6549 | 4.4446 | 119,284 |
| UMICRO | 0.0001 | 0.0000 | 0.0002 | 0.0000 | -0.0003 | 0.0005 | 0.0001 | 0.0003 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0001 | -0.0012 | 0.0036 | -0.0642 | 0.1126 | -0.1126 | 0.2165 | 0.0012 | 2995 |

Table A4. Parameters for glycated haemoglobin function.

| Age (a) | θ | φ | ψ |
|---------|----------|-----------|--------|
| 0–19 | 5.976 | 0.2493 | 0.0107 |
| 20–39 | 3.740 | 0.4810 | 0.0064 |
| 40–64 | 5.626 | 0.1805 | 0.0061 |
| 65–74 | 7.389 | −0.1011 | 0.0057 |
| 75–84 | 7.027 | −0.0611 | 0.0038 |
| 85 + | 7.076 | −0.0741 | 0.0020 |

Appendix B.2. Numerical Experiments

Based on data reported in NHRID, we compute the probability of 5 years’ and 10 years’ complication risks of diabetes according to a nonhomogeneous Markovian process. The probabilities of complication risks of diabetes are generated through the formulae mentioned above, assuming the DM patient starts from 55 years old on average. These possible outcomes by cycles (months) are given in the Supplementary Material. Summarised results are depicted in Figures A1 and A2.

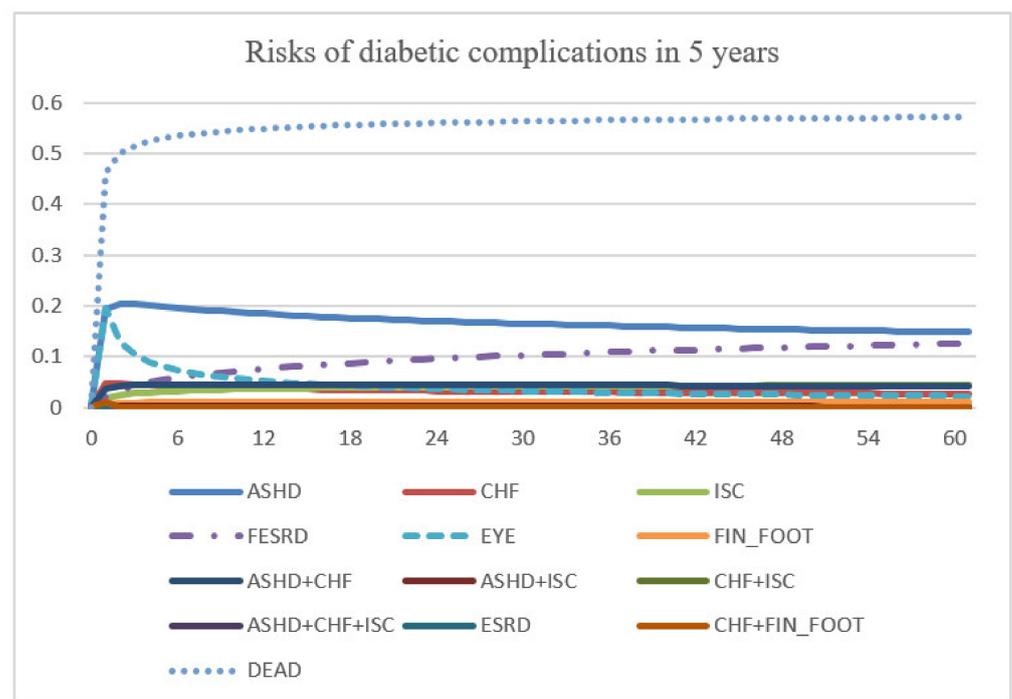


Figure A1. Risks of diabetic complication in 5 years.

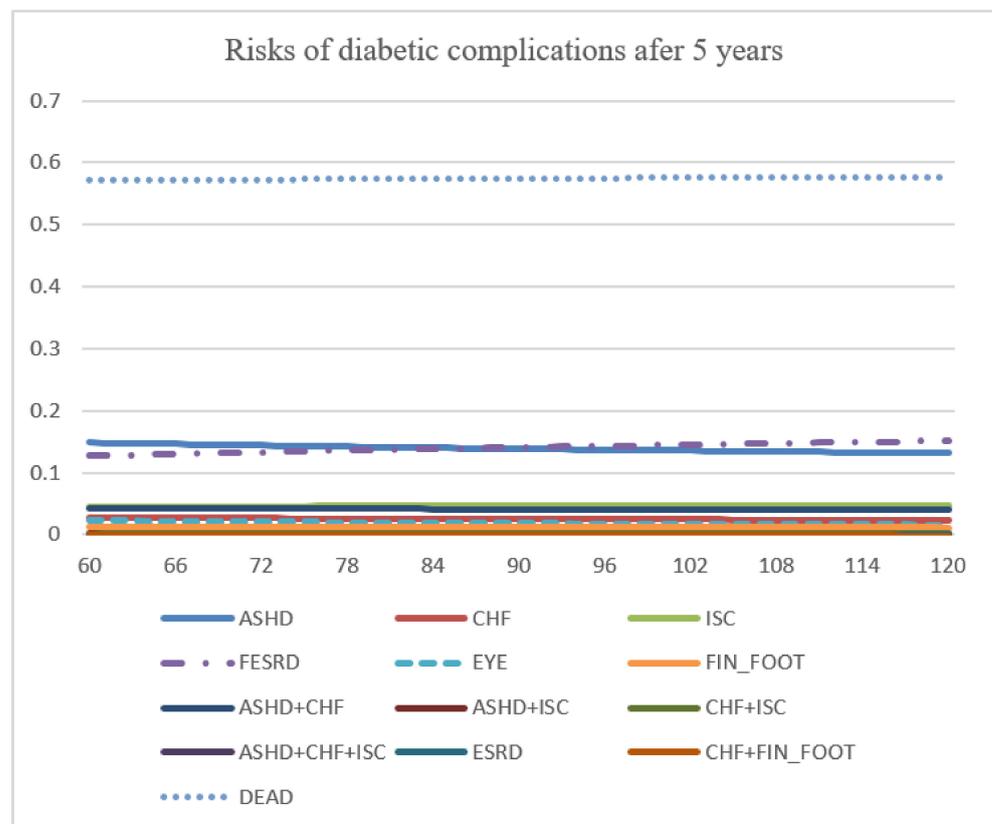


Figure A2. Risks of diabetic complication after 5 years.

References

- International Diabetes Federation (IDF). In *IDF Diabetes Atlas*, 9th ed.; International Diabetes Federation: Brussels, Belgium, 2019.
- Hsieh, C.Y.; Su, C.C.; Shao, S.C.; Sung, S.F.; Lin, S.J.; Yang, Y.H.K.; Lai, E.C. Taiwan's National Health Insurance Research Database: Past and future. *Clin. Epidemiol.* **2019**, *11*, 349–358. [[CrossRef](#)] [[PubMed](#)]
- Lin, L.Y.; Warren-Gash, C.; Smeeth, L.; Chen, P.C. Data resource profile: The National Health Insurance Research Database (NHIRD). *Epidemiol. Health* **2018**, *40*, e2018062. [[CrossRef](#)]
- Lee, P. *2017–2018 National Health Insurance Annual Report*; National Health Insurance Administration, Ministry of Health and Welfare: Taipei, Taiwan, 2017.
- Lee, T.T.; Cheng, S.H.; Chen, C.C.; Lai, M.S. A pay-for-performance program for diabetes care in Taiwan: A preliminary assessment. *Am. J. Manag. Care* **2010**, *16*, 65–69.
- Zheng, Y.; Ley, S.H.; Hu, F.B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat. Rev. Endocrinol.* **2018**, *14*, 88–98. [[CrossRef](#)]
- Wang, C.Y.; Tu, S.T.; Sheu, W.H.; Chen, I.C.; Chuang, L.M.; Wu, M.S.; Yu, C.J. National survey of ABC (A1C, blood pressure, cholesterol) of Diabetes Health Promotion Institutes in Taiwan: 2002–2018. *J. Formos. Med. Assoc.* **2018**, *117*, 952–954. [[CrossRef](#)] [[PubMed](#)]
- Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting Diabetes Mellitus With Machine Learning Techniques. *Front. Genet.* **2018**, *9*, 515. [[CrossRef](#)] [[PubMed](#)]
- Cichosz, S.L.; Johansen, M.D.; Hejlesen, O. Toward Big Data Analytics: Review of Predictive Models in Management of Diabetes and Its Complications. *J. Diabetes Sci. Technol.* **2015**, *10*, 27–34. [[CrossRef](#)]
- Yap, M.H.; Chatwin, K.E.; Ng, C.C.; Abbott, C.A.; Bowling, F.L.; Rajbhandari, S.; Boulton, A.J.M.; Reeves, N.D. A New Mobile Application for Standardizing Diabetic Foot Images. *J. Diabetes Sci. Technol.* **2018**, *12*, 169–173. [[CrossRef](#)] [[PubMed](#)]
- Ellahham, S. Artificial Intelligence: The Future for Diabetes Care. *Am. J. Med.* **2020**, *133*, 895–900. [[CrossRef](#)] [[PubMed](#)]
- Clarke, P.M.; Gray, A.M.; Briggs, A.; Farmer, A.J.; Fenn, P.; Stevens, R.J.; Matthews, D.R.; Stratton, I.M.; Holman, R.R.; UK Prospective Diabetes Study (UKPDS) Group. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* **2004**, *47*, 1747–1759. [[CrossRef](#)]
- Shao, H.; Fonseca, V.; Stoecker, C.; Liu, S.; Shi, L. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). *Pharmacoeconomics* **2018**, *36*, 1125–1134. [[CrossRef](#)] [[PubMed](#)]

14. Mount Hood 4 Modeling Group. Computer modeling of diabetes and its complications: A report on the Fourth Mount Hood Challenge Meeting. *Diabetes Care* **2007**, *30*, 1638–1646. [CrossRef] [PubMed]
15. Mt Hood Diabetes Challenge in Asia 2019, Seoul, Korea, 30 November–1 December 2019. Available online: <https://www.mthooddiabeteschallenge.com/asiameeting> (accessed on 10 January 2021).
16. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. *Diabetes Care* **2004**, *27*, 2262–2265. [CrossRef] [PubMed]
17. Tappenden, P.; Chilcott, J.; Brennan, A.; Squires, H.; Stevenson, M. Whole disease modeling to inform resource allocation decisions in cancer: A methodological framework. *Value Health* **2012**, *15*, 1127–1136. [CrossRef] [PubMed]
18. Esensoy, A.V.; Carter, M.W. Health system modelling for policy development and evaluation: Using qualitative methods to capture the whole-system perspective. *Oper. Res. Health Care* **2015**, *4*, 15–26. [CrossRef]
19. Lind, M.; Oden, A.; Fahlen, M.; Eliasson, B. The true value of HbA1c as a predictor of diabetic complications: Simulations of HbA1c variables. *PLoS ONE* **2009**, *4*, e4412. [CrossRef]
20. Coyle, D.; Lee, K.M.; O'Brien, B.J. The role of models within economic analysis: Focus on type 2 diabetes mellitus. *Pharmacoeconomics* **2002**, *20* (Suppl. 1), 11–19. [CrossRef] [PubMed]
21. Allalou, A.; Nalla, A.; Prentice, K.J.; Liu, Y.; Zhang, M.; Dai, F.F.; Ning, X.; Osborne, L.R.; Cox, B.J.; Gunderson, E.P.; et al. A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes. *Diabetes* **2016**, *65*, 2529–2539. [CrossRef]
22. Tarride, J.E.; Hopkins, R.; Blackhouse, G.; Bowen, J.M.; Bischof, M.; Von Keyserlingk, C.; O'Reilly, D.; Xie, F.; Goeree, R. A review of methods used in long-term cost-effectiveness models of diabetes mellitus treatment. *Pharmacoeconomics* **2010**, *28*, 255–277. [CrossRef]
23. Yi, Y.; Philips, Z.; Bergman, G.; Burslem, K. Economic models in type 2 diabetes. *Curr. Med. Res. Opin.* **2010**, *26*, 2105–2118. [CrossRef]
24. CDC Diabetes Cost-effectiveness Group. Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes. *JAMA* **2002**, *287*, 2542–2551. [CrossRef]
25. Wolbers, M.; Koller, M.T.; Wittman, J.C.; Steyerberg, E.W. Prognostic models with competing risks: Methods and application to coronary risk prediction. *Epidemiology* **2009**, *20*, 555–561. [CrossRef]
26. Soikkeli, F.; Hashim, M.; Ouwens, M.; Postma, M.; Heeg, B. Extrapolating Survival Data Using Historical Trial-Based a Priori Distributions. *Value Health* **2019**, *22*, 1012–1017. [CrossRef]
27. Grossetti, F.; Ieva, F.; Paganoni, A.M. A multi-state approach to patients affected by chronic heart failure: The value added by administrative data. *Health Care Manag. Sci.* **2018**, *21*, 281–291. [CrossRef]
28. Schafer, J.L. Multiple imputation: A primer. *Stat. Methods Med. Res.* **1999**, *8*, 3–15. [CrossRef] [PubMed]
29. Marshall, D.A.; Burgos-Liz, L.; IJzerman, M.J.; Crown, W.; Padula, W.V.; Wong, P.K.; Pasupathy, K.S.; Higashi, M.K.; Osgood, N.D.; the ISPOR Emerging Good Practices Task Force. Selecting a dynamic simulation modeling method for health care delivery research-part 2: Report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. *Value Health* **2015**, *18*, 147–160. [CrossRef]
30. Karnon, J.; Stahl, J.; Brennan, A.; Caro, J.J.; Mar, J.; Moller, J. Modeling using discrete event simulation: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Med. Decis. Mak.* **2012**, *32*, 701–711. [CrossRef]
31. Tanaka, S.; Tanaka, S.; Iimuro, S.; Yamashita, H.; Katayama, S.; Akanuma, Y.; Yamada, N.; Araki, A.; Ito, H.; Sone, H.; et al. Predicting macro- and microvascular complications in type 2 diabetes: The Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care* **2013**, *36*, 1193–1199. [CrossRef] [PubMed]
32. Koo, B.K.; Oh, S.; Kim, Y.J.; Moon, M.K. Prediction of coronary heart disease risk in Korean patients with diabetes mellitus. *J. Lipid Atheroscler.* **2018**, *7*, 110. [CrossRef]
33. Hong Kong Diabetes Registry; Yang, X.; So, W.Y.; Tong, P.C.; Ma, R.C.; Kong, A.P.; Lam, C.W.; Ho, C.S.; Cockram, C.S.; Ko, G.T.; et al. Development and validation of an all-cause mortality risk score in type 2 diabetes. *Arch. Intern. Med.* **2008**, *168*, 451–457. [CrossRef] [PubMed]
34. Yang, X.; So, W.-Y.; Kong, A.P.; Ma, R.C.; Ko, G.T.; Ho, C.-S.; Lam, C.W.; Cockram, C.S.; Chan, J.C.; Tong, P.C. Development and validation of a total coronary heart disease risk score in type 2 diabetes mellitus. *Am. J. Cardiol.* **2008**, *101*, 596–601. [CrossRef] [PubMed]
35. Yang, X.; So, W.; Kong, A.; Clarke, P.; Ho, C.; Lam, C.; Ng, M.; Lyu, R.; Yin, D.; Chow, C. End-stage renal disease risk equations for Hong Kong Chinese patients with type 2 diabetes: Hong Kong Diabetes Registry. *Diabetologia* **2006**, *49*, 2299–2308. [CrossRef] [PubMed]
36. Ma, R.C.; Chan, J.C. Type 2 diabetes in East Asians: Similarities and differences with populations in Europe and the United States. *Ann. N. Y. Acad. Sci.* **2013**, *1281*, 64. [CrossRef] [PubMed]