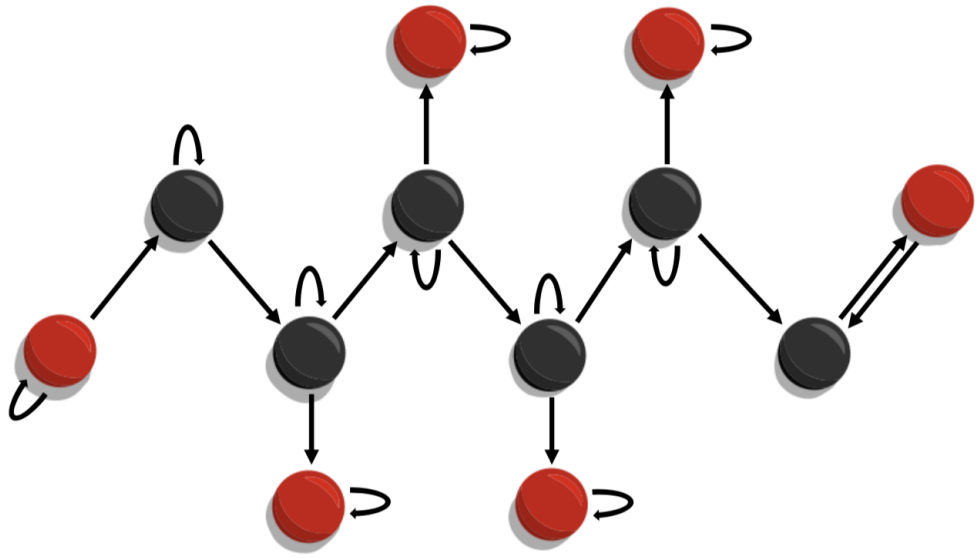
CHALMERS
UNIVERSITY OF TECHNOLOGY



Multi-State Markov Model for Analysing Blood Glucose Changes

A Study Using Continuous Glucose Measurements From Patients With Type 1 Diabetes

Master's thesis in Engineering Mathematics and Computational Science

VIKTOR INGEMARSSON
MARCUS SVENSSON

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020

Multi-State Markov Model for Analysing Blood Glucose Changes

A Study Using Continuous Glucose Measurements From Patients With Type 1 Diabetes

VIKTOR INGEMARSSON, MARCUS SVENSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
Gothenburg, Sweden 2020

Multi-State Markov Model for Analysing Blood Glucose Changes
A Study Using Continuous Glucose Measurements From Patients With Type 1 Diabetes
VIKTOR INGEMARSSON, MARCUS SVENSSON

Supervisors: Ruben Buendia Lopez, Senior Data Scientist at AstraZeneca. Michail Doulis, Senior Data Scientist at AstraZeneca. Jesper Havsol, Informatics Science Director at AstraZeneca.
Examiner: Staffan Nilsson, Associate professor in mathematical statistics, Mathematical Sciences, Chalmers University of Technology

Master's Thesis 2020
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A representation of a mainly progressive Markov model with multiple absorbing states, presented as a molecule similar to glucose.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Multi-State Markov Model for Analysing Blood Glucose Changes
A Study Using Continuous Glucose Measurements From Patients With Type 1 Diabetes

VIKTOR INGEMARSSON, MARCUS SVENSSON
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

A multi-state Markov model was adapted in order to model glycemic control, based on continuous glucose measurements (CGM). A library was implemented, written in Python, that allows for user-specific input in regards to modelling parameters and analysis. The CGM-readings were collected during two previous clinical trials, involving patients with type 1 diabetes and inadequate glycemic control. The clinical trials involved the administration of dapagliflozin which together with insulin improves glycemic control, compared with only administering insulin. The states of the Markov model were defined based on blood glucose levels, where increased time in the target range, normoglycemia, constituted better glycemic control. Based on the CGM-readings collected, this Markov model was used to analyse how the glycemic control of patients is affected by their kidney function as well as insulin reduction. Results show that improvement in glycemic control due to dapagliflozin is independent of kidney function in the range investigated. When modelling the insulin reduction in patients, it was seen that an increased insulin usage corresponded to increased glucose levels. There is a well established causal relationship between insulin and decreased blood glucose levels. The opposite relation seen in the modelling and data must mean that something is masking the effect. One explanation would be that this is due to the fact that insulin is only a proxy for eating unevenly, but the exact cause is unknown.

Keywords: logistic regression, longitudinal data, Markov process, multi-state model,

Acknowledgements

Throughout this thesis we, the writers, have received support and advice from many people, without whom this thesis would never have been written. Firstly, we thank our supervisors Ruben Buendia Lopez, Michail Doulis and Jesper Havsol for their continuous guidence. Lastly we would like to thank AstraZeneca for giving us the opportunity to do this project, and for providing office space and the hardware needed, as well as vast amounts of coffee.

Viktor Ingemarsson and Marcus Svensson, Gothenburg, June 2020

Contents

List of Figures

List of Tables

Acronyms

BFGS Broy-den–Fletcher–Goldfarb–Shanno. 9, 10

CGM Continuous glucose measurement. xi, xiii, 2, 4, 6, 15, 16, 17, 18, 20, 21, 23

DKA Diabetic ketoacidosis. 1, 4

GFR Glomerular filtration rat. 16
GRF Self monitoring of blood glucose. 6

MST Mean sojourn time. 7

SDAM Squared deviations for array mean. 11
SGLT2 Sodium-glucose co-transporter. 2
SSDCM Sum of squared deviations for class means. 11

T1DM Type 1 diabetes. xi, 1, 2, 5
T2DM Type 2 diabetes. 1, 2
TPM Transition probability matrix. 6, 7, 13

1
Introduction

When eating, the human body will break down the carbohydrates in the food, which are then distributed through the bloodstream to all cells. These carbohydrates are then absorbed and used as energy. Glucose, a subcategory of carbohydrates, in the bloodstream can be discretised into categories depending on its levels, where a target range, known as normoglycemia, has been identified. Levels above this target range means a patient is in hyperglycemia and levels below target range is referred to as hypoglycemia.

A healthy body regulates its glucose levels by the pancreas producing a sufficient amount of a hormone called insulin. Insulin promotes the absorption of carbohydrates, especially glucose, into the cells which decreases the glucose levels in the blood again after an increase from eating. Prolonged periods of hyperglycemia have been linked to heart disease [1, 2, 3], stroke [4, 5], kidney disease [6], vision problems [7], and nerve problems [8, 1]. Hyperglycemia is also commonly seen in the case of the serious acute complication of diabetic ketoacidosis (DKA) which is caused by insulin doses or levels that are much lower than that needed. The response from the body is to switch to burning fatty acids for energy which produces ketone bodies lowering the pH of the blood which, if left untreated, leads to coma and death. Conversely, an excess of insulin might lower the amount of glucose in the blood to critical levels, leading to hypoglycemia. This state may lead to nausea, headaches, unconsciousness and, in occasional cases when not treated, death.

Diabetes mellitus, commonly known as diabetes, is a group of metabolic diseases in which the body lose part of its ability to regulate glucose levels. There are mainly two kinds of diabetes: T1DM and type 2 diabetes (T2DM). T2DM is characterised as a combination of failure by the body to produce enough insulin and the cells growing resistant to insulin leading to risk of entering hyperglycemia and other complications of diabetes. People suffering from T1DM lose all, or almost all, ability to produce insulin. Because of this, the body cannot transfer energy to cells efficiently which leads to that these individuals risk becoming hyperglycemic. To mitigate the risk of entering a hyperglycemic state, patients suffering from these conditions continuously monitor their glucose levels and self-administer insulin through injections to regulate glucose levels. A diabetic patient's ability to adapt to changes in blood glucose, mitigating the risk of entering hypoglycemia or hyperglycemia and instead increase the portion of stay in normoglycemia is called glycemic control. Regulation of glycemia through exogenously administered insulin is challenging and frequently leads to fluctuations in glucose. These fluctuations may in themselves be associated

with complications of diabetes such as endothelial dysfunction and atherosclerosis [9]. Moreover, this self administration, even though improved with new technology, is suboptimal in relation to keeping steady levels of glucose. Patients with this set of diseases not only have greater fluctuations in glucose but also higher average levels than those without the diseases. Comparison of prevalence of glucose readings between T1DM patients and patients not suffering from any glycemic control diseases can be seen in Figure 1.1. Data for the non-diabetics has been taken from a small previous study with public data [10] and data for the T1DM diabetics are from two phase three studies called DEPICT [11, 12].

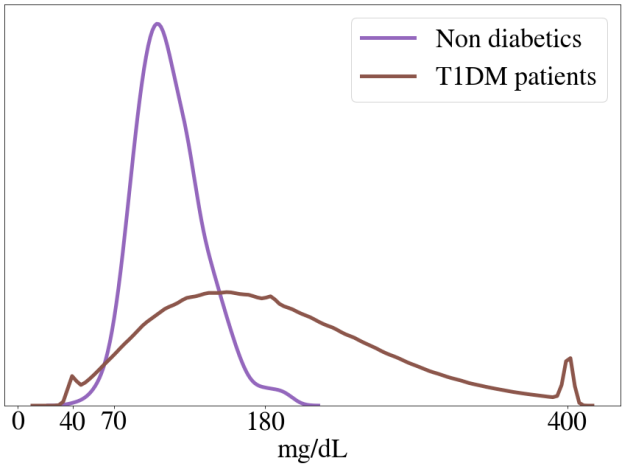


Figure 1.1: Comparison of probability density curves of readings for T1DM patients treated only with insulin, taken from the DEPICT-trials, and non diabetics, from a small sample in an old study.

Recently a new group of medicines, which have a proven positive effect on the glycemic control in T2DM patients, have been developed. Medicines in this group are called sodium-glucose co-transporter (SGLT2) inhibitors. They work by inhibiting sodium-glucose transport proteins to reabsorb glucose into the bloodstream when passing the kidneys, and instead removing it from the body via urine. This stabilises the levels of glucose and decreases the need for insulin. One such inhibitor, developed by AstraZeneca, is called dapagliflozin. This is an approved medicine for treating T2DM patients in both in many regions, including the US and EU and has shown positive effects on cardiovascular disease, heart failure and progression of renal disease[13, 14]. Two large randomised, placebo-controlled clinical trials have been carried out, where patients suffering from T1DM have tested two doses of dapagliflozin together with insulin. Data from these two clinical trials have been analysed using traditional statistical tests and methods, which show that patients on dapagliflozin have a more favorable glycemic control based on their levels of glycated hemoglobin (HbA1c). HbA1c is an indicator of the average level of blood sugar over the past 3 months, and is tested weeks apart as a proxy for glycemic control. During parts of the DEPICT-trials, the patients had CGM devices, which allowed for precise measurements of glucose levels in their blood. Given this collected CGM-data

of high granularity, further analysis into the glycemic control of the patients is made possible, and can serve as a complementary tool to the 3 month average, HbA1c.

Since dapagliflozin, just like insulin, lowers glucose levels, patients were asked to reduce their insulin dosages appropriately, with the purpose of not entering hypoglycemic state unexpectedly. Patients were then asked to attempt to titrate back up again. Resuming the earlier dosages was as expected more easily achieved for those patients that received placebo, leading to greater dose reductions in those receiving dapagliflozin. This may partially offset the glycemic effect of dapagliflozin in the studies. There is a sought after need to quantify how much different levels of insulin reduction affect glycemic control. Furthermore, a model that models glycemic control whilst incorporating many different features such as insulin usage as well as age, gender, body mass index etc. then becomes a crucial step in order to determine how much any of these different aspects affect the level of glycemic control different patient groups have.

A widely used approach to model transition between states, within diseases as well as other disciplines such as finance, is the multi-state Markov model [15, 16, 17, 18, 19, 20, 21, 22]. This Markov model models the states of the system, in this case the glycemic states, with variables that change through time. This is a highly flexible method that can handle time-inhomogeneous and time-homogeneous processes. Moreover, Markov models can incorporate the effects of features by either discretisation of data dependent on the feature thus creating multiple models, or by application of regression on the transition probabilities between the predefined states. A simple Markov model inhibits the Markov property, which asserts that the future states of the process depend only on the present state, not on the sequence of events that preceded it.

1.1 Aim

This project aims to investigate the suitability to model glucose changes using a Markov model, as well as developing such a model. By implementing this model using the DEPICT-data, this project also aims to find and analyse what relation different features, other than treatment arm and insulin, have on glycemic control. Following this analysis, evaluation of the relation between insulin dosage and glycemic control, for the different treatment arms will be performed.

1.2 Research Questions

- How should the Markov model be adapted and implemented in this setting to draw insights?
- How do features, other than the usage of dapagliflozin and insulin, relate to glycemic control?
- How do changes in dosage of insulin, when combined with dapagliflozin treatment, affect glycemic control?

1.3 Scope

The model will be made specifically for the DEPICT studies and no effort will be made to make it more general than necessary to fit any other data set. Analysis will be done only on these two studies. Even though the studies include severe events, such as DKA, these will not be included in the model because of its low prevalence in the data which would be hard to model.

This project aims at developing a model that models glycemic control. The glucose levels during the day, as recorded by the CGM device, mainly depend on when you eat, what you eat, as well as when and how much insulin you take. Data regarding this from the DEPICT studies is somewhat limited. Data in regards to physical exercises is not collected. Insulin doses are recorded as daily measurements with no time stamps on when insulin was injected. Meal-times are recorded as diary-entries with time stamps for breakfast, lunch and dinner. There exists no data on what is consumed, neither are there any entries for any other forms of consumption, such as snacking or drinking. Because of this, a model that accurately models intra-day glycemia levels can never be created, since the major drivers, as explained above, are lacking granularity. What is hoped to be achieved with the Markov model however, is to distinguish the effect of attributes on a higher level, such as patient specific features and treatment arms.

2
Theory

Information in regards to the setup of two studies, also known as DEPICT-studies, and which records are used is presented, though heavily reduced. For a comprehensive review of the studies readers are referred to previous publications in the two articles *Efficacy and safety of dapagliflozin in patients with inadequately controlled type 1 diabetes (DEPICT-1): 24 week results from a multicentre, double-blind, phase 3, randomised controlled trial* [12] and *Efficacy and Safety of Dapagliflozin in Patients With Inadequately Controlled Type 1 Diabetes (the DEPICT-2 Study): 24-Week Results From a Randomised Controlled Trial* [11]. Theory behind the methods used is presented in its general form and readers with knowledge in the field are referred to the methodology section on how general methods were applied to the given set of problems.

2.1 The DEPICT-Studies

The two studies which contents have been used in the analysis have both been 24 week long, three legged, double-blind, phase three, randomised controlled trials. In the trials, patients with T1DM were randomised in ratio of 1:1:1 into each leg of the studies, 5 or 10 mg dapagliflozin per day or placebo, where each patient group complemented the administration of dapagliflozin/placebo with insulin adjusted as deemed appropriate. Patients were in all ages ranging from 18 to 75 and all had been prescribed and used insulin for a minimum of 12 months. The studies were conduced in the following countries: Argentina, Australia, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Hungary, Israel, Italy, Japan, Mexico, the Netherlands, Poland, Romania, Russia, Spain, Sweden, Switzerland, U.K and the U.S. Both studies have been compliant with the Declaration of Helsinki and Good Clinical Practice Guidelines as defined by the International Conference on Harmonisation. Both of which are guidelines and principles on how human experiments should be performed and documented to ensure safety of participants, but also harmonisation in results between studies. Patients eligible to participate in the studies entered an eight week lead-in period where patients' glucose values were registered, diet and exercise guidance was given and insulin was optimised. This gives insight into individual glycemic control and the possibility to assess variability in glycemic profiles within the study participants together with a baseline of insulin usage on an individual level. After the lead-in period was finished participants entered a 24 week long treatment period medicating dependent on treatment arm, each given once per day. Data from the studies was registered in three separate indices, last

two weeks of lead-in period, week 10-12 and 22-24 in treatment period. Each of these two week periods will from now on be referred to as the lead-in period, treatment period one and treatment period two. Initially, patients were recommended to reduce insulin by up to 20 % when the study phase was initiated and then urged to attempt to titrate insulin doses back to baseline levels. Insulin dosages consists of two parts: basal and bolus. Basal, also known as background insulin, can be one of two things: either a long acting insulin taking effect over the whole day or a constant infusion rate of insulin, delivered through a catheter from a mechanical pump. Both methods have the purpose of bringing down high resting glucose levels during periods of fasting. Bolus is instead taken at times of food consumption or moments of high glucose levels to bring down glucose levels to appropriate levels rapidly. Recommendation to reduce insulin was issued to minimise the risk of entering hypoglycemic state when administered blinded study medication. In addition to the patients' usual self monitoring of blood glucose (SMBG) or CGM readings a study-CGM was recorded. The monitor Dexcom G4 Platinum was used to collect the data for the analysis. Patients were also urged to record mealtimes and their administration of medication.

2.2 Multi-State Markov Model

When observing a change in a process, and such change is between pre-defined states, a multi-state Markov model can be used. Observations are often recorded with some time in between, thus creating uncertainty of what happens between observations. If observations are recorded with sufficiently small time differences, the uncertainties become small, and the exact time of the state change can be recorded. Given such observations, a transition probability matrix (TPM), $\mathbf{P}$, can be defined as

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}, \tag{2.1}$$

where n is the number of pre-defined states. Moreover, p_{ij} is the probability of changing from state i to state j, where $i,\ j \in \{1, ..., n\}$. The probabilities of going from any one state to any other state sums to one, which means that each row in the TPM $\mathbf{P}$ sums to one. Clearly, the probabilities for state changes that cannot take place become 0.

The TPM $\mathbf{P}$ might not be static, but instead depend on features. Each transition probability then becomes dependent on said features. The TPM then becomes

$$\mathbf{P} = \begin{bmatrix} p_{11}(\boldsymbol{x}) & \cdots & p_{1n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ p_{n1}(\boldsymbol{x}) & \cdots & p_{nn}(\boldsymbol{x}) \end{bmatrix}, \tag{2.2}$$

where $p_{ij}(\boldsymbol{x})$ is a function describing the probability transitioning from state i to state j given the covariates $\boldsymbol{x}$. If the observations are recorded with a sufficiently

small time difference, in combination with that the states are sequential, transitions are only possible between neighbouring states. The TPM then becomes

$$\mathbf{P} = \begin{bmatrix} p_{11}(\boldsymbol{x}) & p_{12}(\boldsymbol{x}) & 0 & 0 & \dots & 0 & 0 & 0 \\ p_{21}(\boldsymbol{x}) & p_{22}(\boldsymbol{x}) & p_{23}(\boldsymbol{x}) & 0 & \dots & 0 & 0 & 0 \\ 0 & p_{32}(\boldsymbol{x}) & p_{33}(\boldsymbol{x}) & p_{34}(\boldsymbol{x}) & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & p_{nn-1}(\boldsymbol{x}) & p_{nn}(\boldsymbol{x}) \end{bmatrix}, \tag{2.3}$$

2.2.1 Mean Sojourn Time

Mean sojourn time (MST) is the time an object is expected to stay within any system before exiting again. Calculation of this measure can be done by summing up the amount of time spent within any system divided by the number of times exiting the system. Together with the TPM defined in Equation 2.3, the mean sujourn time becomes

$$\begin{aligned} MST_1 &= \frac{1}{p_{12}(\boldsymbol{x})}\Delta t \\ MST_i &= \frac{1}{p_{i,i-1}(\boldsymbol{x}) + p_{i,i+1}(\boldsymbol{x})}\Delta t \qquad i \in \mathcal{Z} : i \in [2, n-1] \\ MST_n &= \frac{1}{p_{n,n-1}(\boldsymbol{x})}\Delta t \end{aligned}$$

where n is the total number of states and Δt is the time between each time step.

2.2.2 Fraction of Time

Let $\mathbf{P}$ be the TPM as defined in Equation 2.3. Let $\mathbf{u}$ be the vector which represents the starting probability in each state, where $\sum_i^n u_i = 1$ where n is the number of states. Then the probability that the Markov model is in state s_i after t time-steps is the i th entry in the vector

$$\mathbf{u}(t) = \mathbf{u}(\mathbf{t}-\mathbf{1})\mathbf{P}^t. \tag{2.4}$$

The probability of being in a certain state, s_i, is also the fraction of time spent in that state given the TPM, $\mathbf{P}$. A vector of extra interest is the probability vector $\mathbf{u}$ such that

$$\mathbf{u} = \mathbf{u}\mathbf{P}. \tag{2.5}$$

This vector is of extra interest since when a steady state is reached, multiplication of vector $\mathbf{u}$ by $\mathbf{P}$ does not affect the resulting vector $\mathbf{u}$. This is the probability of being in any of the possible states, i, given the TPM after infinite time which can be interpreted as the portion of time being in each state. To get this vector $\mathbf{u}$, one can use the eigendecoposition of matrices. Since each row of TPM sums to one then

$$det(\mathbf{P} - 1 \cdot \mathbf{I}) = 0.$$

Given that the eigenvalue decomposition

$$\mathbf{Pu} = \lambda\mathbf{u} \tag{2.6}$$

and solution to λ is given as

$$det(P - \lambda I) = 0$$

then one eigenvalue corresponding to the TPM must be 1. Given that one eigenvalue is one the corresponding steady state vector, an eigenvector of $\mathbf{P}$, is given by solving Equation 2.6.

2.3 Logistic Regression

In order to predict outcomes of a categorical nature, either binomial or multinomial logistic regression can be used. For the binary outcome case, the probability of outcome one, p, given some observation $\boldsymbol{x}$ is

$$p(\boldsymbol{x}) = \frac{1}{1 + e^{-\beta^T \boldsymbol{x}}}, \tag{2.7}$$

where $\boldsymbol{\beta}$ is a vector of coefficients. The probability of outcome two, clearly becomes $1 - p$.

In order to find the parameters $\boldsymbol{\beta}$ that gives the most accurate predictions, they have to be estimated from the data points. Unlike linear regression, where the optimal parameters can be computed in closed form, the normal equation, logistic regression requires a different approach. In logistic regression, the logistic loss function is minimised in an iterative process. This cost is defined as

$$Cost(p(\boldsymbol{x}), y) = \begin{cases} -log(p(\boldsymbol{x})) & \text{if } y = 1 \\ -log(1 - p(\boldsymbol{x})) & \text{if } y = 0 \end{cases}. \tag{2.8}$$

Although two cases are studied, when y is either 0 or 1, the cost can be rewritten as

$$Cost(p(\boldsymbol{x}), y) = -ylog(p(\boldsymbol{x})) - (1 - y)log(1 - p(\boldsymbol{x})). \tag{2.9}$$

The cost function, $J(\boldsymbol{\beta})$, of the model then becomes the summation from all the data points used

$$J(\boldsymbol{\beta}) = \frac{1}{m} \sum_{k=1}^{m} Cost(p_{\beta}(\boldsymbol{x}^{(k)}), y^{(k)}) \tag{2.10}$$

which becomes

$$J(\boldsymbol{\beta}) = -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} log(p(\boldsymbol{x}^{(i)})) + (1 - y^{(i)}) log(1 - p(\boldsymbol{x}^{(i)})) \right), \tag{2.11}$$

where m is the number of samples. This cost function seen in Equation 2.11, is convex, which makes the optimisation used to find a global optima highly favorable [23].

For the multinomial case, where there are K possible outcomes, the probabilities are instead computed using softmax regression, thus becoming

$$p_i(\mathbf{x}) = \frac{e^{\boldsymbol{\beta}_i \cdot \mathbf{x}}}{\sum_{k=1}^{K} e^{\boldsymbol{\beta}_k \cdot \mathbf{x}}} \quad i \in \mathcal{Z} : i \in [1, K] \tag{2.12}$$

where p_1 denotes the probability of outcome when $K = 1$. Since the denominator sums to one, it has a normalising effect, thus allowing for the sum of probabilities to become one. As can be seen, the vector of coefficients, $\boldsymbol{\beta}$, vary depending on what outcomes are being predicted. The cost function, $J(\boldsymbol{\beta})$, in the multinomial case is instead computed as seen below in Equation 2.13,

$$J(\boldsymbol{\beta}) = -\sum_{i=1}^{m}\sum_{k=1}^{K}\left(1\{y^{(i)} = k\} log\Big(P(y^{(i)} = k|x^{(i)}; \boldsymbol{\beta})\Big)\right). \tag{2.13}$$

These cost functions, both for the binomial as well as the multinomial case as described in Equation 2.11 and 2.13 respectively, are to be minimised in order to find the set of parameters that minimises the cost. This in turn maximises the predictive power of the logistic regression model. In order to perform such minimisation of cost, an optimisation algorithm can be used. One such algorithm is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, which belongs to quasi-Newton methods, for the numerical optimisation of any function, $f(\theta)$, with respect to its parameters θ. In the particular case of logistic regression, $J(\boldsymbol{\beta})$ can be optimised with respect to its parameters $\boldsymbol{\beta}$. In contrast to Newton's Method, the Hessian matrix of second derivatives is not computed for Quasi-Newton methods. Approximation is instead used to establish the Hessian matrix, using approximate gradient evaluations. The BFGS algorithm uses a modified way of updating the approximate Hessian matrix, which is described in more detail below. An initial guess is made, usually I, for an approximate Hessian matrix B_0 as well as an initial guess for the variable $\boldsymbol{\beta}_0$. The following steps are repeated as $\boldsymbol{\beta}_k$ converges to a solution. Here, the indice k denotes the iterations performed by the BFGS algorithm.

1. Compute a search direction $\boldsymbol{p}_k$ by solving $B_k\boldsymbol{p}_k = -\nabla f(\boldsymbol{\beta}_k)$.
2. Find a step size α_k by doing a line search in the direction of the search, so $\alpha_k = \arg\min f(\boldsymbol{\beta}_k + \alpha_k\boldsymbol{p}_k)$.
3. Define $\boldsymbol{s}_k = \alpha_k\boldsymbol{p}_k$ and update the parameters as $\boldsymbol{\beta}_{k+1} = \boldsymbol{\beta}_k + \boldsymbol{s}_k$.
4. $\boldsymbol{y}_k = \nabla f(\boldsymbol{\beta}_{k+1}) - \nabla f(\boldsymbol{\beta}_k)$.
5. Update the approximate Hessian as $B_{k+1} = B_k + \dfrac{\mathbf{y}_k\mathbf{y}_k^{\mathrm{T}}}{\mathbf{y}_k^{\mathrm{T}}\mathbf{s}_k} - \dfrac{B_k\mathbf{s}_k\mathbf{s}_k^{\mathrm{T}}B_k^{\mathrm{T}}}{\mathbf{s}_k^{\mathrm{T}}B_k\mathbf{s}_k}$.

When the norm of the gradient, $||\nabla f(\boldsymbol{\beta}_k)||$, is sufficiently small or a maximum number of iterations are completed, the algorithm stops. Convergence to a global optima cannot be guaranteed, since the BFGS approximation may not converge to the true Hessian matrix. The final coefficients reached when the iterations stop, $\boldsymbol{\beta}^*$, is used as the coefficients for when constructing the logistic regression model. An improved version of the algorithm, which improves performance when performed on a computer, is the limited-memory BFGS (L-BFGS). The numerical approximations are computed the same way, but old gradients are discarded to leave more space

for freshly computed gradients, which makes it require less memory, which can be beneficial for a computer setting.

2.4 Data Manipulation Methods

In order to perform certain modelling, some different manipulation methods need to be applied to the data. This is done in order to allow for certain algorithms to converge, as well as allowing for certain features to be included.

2.4.1 Feature Scaling

The method of normalising the range of independent variables is called feature scaling. Ranges in raw data, for different features, often has large discrepancies. A large set of machine learning algorithms depend on normalisation to work properly. Also the difficulty of interpreting result might heavily dependant on normalisation in pre-processing steps. In logistic regression, interpretation of regression coefficients is highly sensitive to the scale of the input. In many classifiers euclidean distances are used which in turn results in features with relative large ranges will govern the results. Also in tests L-BFGS has been shown to be greatly accelerated, reaching convergence faster, when normalisation is applied [24].

2.4.2 Dummy-Variables

A feature is said to be categorical if it is not of numeric nature, or that it has no logical ordering of its values. Because of this, such a feature takes on one of a limited, pre-specified, number of values. These are referred to as the levels of the categorical feature. In order to include categorical features when the parameter estimation of the Markov model is performed, they have to be modified. This is done by performing one-hot-encoding, which makes each level of the categorical feature an individual, binary, feature, thus creating what is called dummy-variables. For each observation within the data-set, only one of a categorical feature's levels can be true, thus setting it to one, leaving all other binary values to zero. Each level of the feature now corresponds to its own parameter β, which indicates how much the prevalence of that specific level contributes to the outcome.

2.4.3 Interaction Terms

When statistical modelling is performed using some features for inference, the interaction effect between some of the features could be of interest. It could be that the effect of one feature and the outcome depend on the value of another feature, that is, the effects of the two features are not additive. These interaction effects can be captured by including what is called interaction terms. These are computed by multiplying the features with each other, thus creating an additional feature, the interaction term. This term has its own parameter β, which is estimated the same way as the others. This way, the interaction effects of interest are captured within the model, thus allowing for more extensive inference.

2.5 Clustering

Clustering is the art of grouping similar entries into a group or cluster dependent on their attributes. This can be done when it is known that underlying groups exists within a population but its compositions are unknown. Clustering can be done using multiple algorithms in large feature space. One special case of clustering is in one dimensional feature space. Doing clustering in one dimension is special in the way that values can be definitely sorted dependent on its values, this makes the task easier and there is, dependent on the measurable, a definitive best solution. Jenks natural breaks optimisation solves this problem by calculating the best ranges for a set of data given the number of clusters. It does this by first calculating the sum of squared deviations for array mean (SDAM),

$$SDAM = \sum_{x \in X} (\bar{x} - x_i)^2, \tag{2.14}$$

where X is the set of all values and $\bar{x}$ is the mean. Then split the data into all possible combinations of clusters which will depend on the number of clusters chosen to be present in the data. The algorithm will solve the problem naively, testing all the possible splits, which means that computation-wise it will depend heavily on the number of possible splits. The number of ways it is possible to split the data of size n in k non-empty clusters is known as the Stirling partition number and is calculated as

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n.$$

Calculate the sum of squared deviations for class means (SSDCM) for each of the possible split, j,

$$SSDCM_j = \sum_{g \in G} \sum_{x \in X^{(g)}} (x_i^{(g)} - \bar{x}^{(g)})^2, \tag{2.15}$$

where g is one split of all possible ones and $X^{(g)}$ is the set of observations in that split and group. Then calculate the score which is

$$Score = \frac{(SDAM - SSDCM)}{SDAM},$$

the split with the highest score is the best split. This algorithm in pseudo code can be found in Appendix B.

2.6 Validation

When developing multiple models, of any kind, validation can be useful for determining model performance. It is necessary that the model reflects the underlying data adequately.

2.6.1 Brier Score

Brier score is used to evaluate the accuracy models where the output is probability predictions. The sum of predictions over each outcome sums to one and each

predicted outcome is in the range of 0 to 1,

$$1 = \sum_{i=1}^{n} p_i \quad p \in [0, 1]$$

where n is the number of possible outcomes from the model. The outcome of a prediction has to be either binary or categorical. Brier score is calculated as the mean squared difference between the actual probability compared to the predicted probability for an outcome,

$$BS^{(s)} = \frac{1}{N} \sum_{i=1}^{n} (o_i - p_i)^2.$$

This means a lower Brier score, BS, indicates a better calibrated model and also that the score is in the range of 0 to 1, $BS \in [0, 1]$. In applications where multiple categories are considered the reformulation of the score is

$$BS^{(m)} = \frac{1}{N} \sum_{i=1}^{n} \sum_{c \in C} (o_{ic} - p_{ic})^2 \tag{2.16}$$

where C is the different categories. A formulation not mentioned earlier and here proposed is where multiple classifer models are to be compared, the formulation in Equation 2.16 is used but now averaged over all models,

$$BS^{(p)} = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N_m} \sum_{i=1}^{N} \sum_{c \in C_m} (o_{ic} - p_{ic})^2$$

where M is the number of models.

2.6.2 K-Fold Cross Validation

Given an evaluation metric, K-fold cross validation is used to evaluate the model performance on a given data set. The data is split in K parts, where the model is is fitted on $K-1$ parts, and evaluated on the remaining singular part. This is repeated K times, or folds, in a structured manner so that each data point is evaluated upon exactly once. The evaluation metric is then averaged over the K folds to get the overall model performance. This way, the variance of the estimated performance of the model becomes small when K increases, thus allowing for reliable evaluation.

2.6.3 Bootstrapping

When random sampling with replacement is performed, a metric can be estimated over and over again in order to compute its value and variance. This can in turn be used to compute confidence intervals for the sample estimate for the underlying distribution. In the case of estimating the distribution of the sample mean, where the data consists of n observation, random sampling with replacement is performed m times. This results in estimations of the mean, μ_i where $i \in [1, m]$. The unbiased sample variance is computed as

$$s^2 = \frac{1}{m-1} \sum_{i}^{m} (\mu_i - \bar{\mu})^2 \tag{2.17}$$

where $\bar{\mu}$ is computed as $\frac{1}{m}\sum_i^m \mu_i$. The sample standard deviation then becomes, $s = \sqrt{s^2}$.

2.7 Related Work

Multi-state Markov models in continuous time are often used to model the course of diseases in previous literature. This progression of diseases usually have an absorbing state where the patient's disease progress towards, often death. Recovery may at times occur before patients die, allowing for state changes in both directions. In one paper, data on patients with liver cancer, hepatocellular carcinoma, was collected over a period of time and then modelled with a multi-state Markov model [15]. Around two thirds of the patients eventually die because of the disease, thus reaching the final absorbing state, death. In the paper, the authors fit both 3-state and 4-state models, where the disease progression was analysed. Multiple examples of similar analysis of disease progression in continuous time are found. Without going into detail of the specifics of the Markov modelling performed, examples of other related work where such an analysis is performed are: in regards to diabetic nephropathy [17], chronic post-transplant problems [25], human immunodeficiency virus infection [16, 18], diabetic retinopathy [20] and screening for breast cancer [19], all of which has progressive stages of their respective diseases.

As shown, there exists an extensive literature repertoire regarding Markov modelling in a progressive disease setting. However, this project aims at developing a Markov model where it is not the progression of a disease that is to be modeled. This differs in the way that there is no progression towards a finite, absorbing state, eg. death, as is for the previous works mentioned above. Instead, the states change to and from a "mean"-state when time progresses, with no absorbing state. This methodology of states have previously been used outside of clinical data, more specifically in the case of economic forecasting and financial modelling. There exists related work where the states are instead states of economic recession and its opposite, booming [26]. Here, there is no absorbing state, but it is instead assumed that there exists a "mean"-state of economic prosperity, and that the states of the economy change around this. The paper creates a Markov model that, given a set of leading indicators, models the probability of entering and exiting a high financial stress regime. The probabilities of the TPM of the Markov model that is developed are estimated using a logistic regression framework.

3
Methods

Process of the work has been separated into two mainly serial assignments: preprocessing and analysis. Initial first phase included cleaning and enhancement, transformation and feature extraction. The analysis part was not limited to an analysis performed once with data using selected features but included the construction of a flexible library in Python specifically developed for the assignment. The developed library for the task is able to create Markov models and perform model assessments as well as outputting metrics. The model approach is flexible with respect to how states are defined. In previous work, three states are usually defined with ranges in glucose readings, these states are: hypoglycemia (<70mg/dL), normoglycemia (70-180mg/dL) and hyperglycemia (>180mg/dL). In most parts of this project, these predefined states were used.

3.1 Data Characteristics

All data handled in the project was from two independent triple arm double blind placebo trials. Data from each of the two studies were concluded into four sets of data. The sets all contained data needed in the analysis, the sets corresponded to CGM data, patient data, meal times data and medication data, where the medication data consisted of both dapagliflozin dosages and insulin dosages. Sets also corresponded to granularity, see Table 3.1. Sets could be joined on patient identifier, dates and time of measurement.

Table 3.1: The granularity for each of the four data sets from each study.

Data set	CGM	Medication	Meal	Patient
Interval	5 minutes	Daily	Daily	Once

The CGM set contained glucose readings with five minute intervals identified with patient ID together with date and time of the specific reading. According to the DEPICT-studies, the data should have been divided into three separate periods, each consisting of two weeks. One *lead-in* period, where all data is recorded prior to the introduction of any treatment, placebo or any of the two arms with dapagliflozin. Following this period was two periods, each consisting of two weeks, with readings when treatment had been introduced. Recordings were at week 11-12 and 23-24, measured as the number of weeks after the introduction of the treatment. Participants of the study were urged to record their daily intravenous administration of

insulin. Insulin recordings were to be labeled bolus or basal dependent on which type of insulin it was. Moreover, the drug administration of the treatment drug was to be registered, which was done during both the *treatment* periods. Patients in the trials were urged to record the date and time for the three main courses (breakfast, lunch and dinner) of the day in diary format. Recordings were from the last week in each of the two *treatment* periods, week 12 and 24. Patient data had information in regards to patient specific attributes such as age, gender, and also indication on their kidney function, glomerular filtration rate (GFR). It should be noted that GFR is declining with age and a GFR value in the range 60 to 130 is often seen as normal in a healthy population [27, 28]. Patient data also had dates for when patients entered the actual treatment leg. Patient data also had dates for when each recording in the two treatment periods were initiated which should concur with the labeling in the CGM data.

Table 3.2: Data completeness in fractions for each set of data in each week it should be present. Worth noting is that meal data is only collected in one of the two weeks in each period. CGM data completeness is calculated by grouping by patient and period and checking the registered values and dividing this with the expected number, $14 \cdot 24 \cdot 12 = 4032$. The insulin data is grouped on patient ID and date to see if 14 readings exist per period, where the discrepancy from 14 is averaged and presented below. Meal data completeness is calculated by grouping by patient and period and then checking the number of each meal type and divided by 7. The mean of all patients is then calculated.

Period Week	Lead-in -2 & -1	Treatment 11 & 12	Treatment 23 & 24
CGM data	0.79	0.82	0.82
Insulin data	0.87	0.85	0.89
Meal data	0.90	0.92	0.92

3.2 Pre-Processing

The data have been used in previous analysis and should thus have had proper formatting and been labeled correctly. Notably, this assumption was not correct since thorough exploration revealed shortcomings in previous work which could have led to misleading, or even incorrect, results and conclusions. Below, the details regarding the thorough exploration will be presented, and actions performed to correct it will be outlined.

3.2.1 Cleaning and Enhancement

The first step, cleaning and enhancement, includes the process of unveiling corrupt data, incorrectly labeled data, technical shortcomings of instruments used and related causes of bad data. This is important to ensure, so that conclusions that are

drawn are not based on errors in the data.

Looking at the CGM-data it can be noted that measurements above approximately 400 were rare, accounting for less than 0.15 ‰ of the total number of observations. Furthermore, there is a significant peak in the distribution at approximately 400 $\frac{mg}{dL}$ and at 40 $\frac{mg}{dL}$, see Figure 3.1.

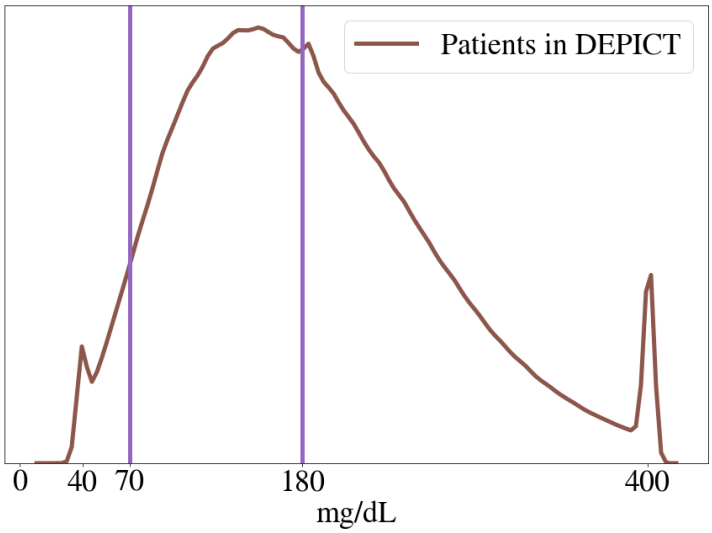


Figure 3.1: Probability density curves of readings of CGM from patients enrolled in the DEPICT-1 and DEPICT-2 studies. Note the high prevalence of readings in hyperglycemia, more than 180 mg/dL of glucose, and low prevalence of observations in hypoglycemia, less than 70 mg/dL of glucose.

The cause of this was the meter being used, Dexcom G4 Platinum, which has a measuring range of 40 - 400 $\frac{mg}{dL}$ [29]. Knowing that the meter only measures within that range, all measurements outside of the range are deemed errors and are removed. These constituted 0.2 ‰ of all values. It also meant that states should not be defined outside of this range.

As previously mentioned, the different sets of data, CGM-data, patient data and medication data had different granularity and were thus separated by this, while joining them was possible on the basis of two features, patient ID's and dates. Looking at the data from DEPICT-1, each CGM reading had labels consisting of *lead-in*, *treatment* or *follow-up*. The patient data on the other hand had one column that stated the first date for an individual receiving any treatment, while there in the medication data was entries on what medication was taken each day. The data in the vast majority of patients contradicted itself on this point. Readings from the CGM-data, labeled *treatment*, were in these cases prior to the first day of treatment in the patient data and in the medicine data, consisting of daily recordings, treatment had not been administered. No information was possible to get on what was correct thus the decision was made to assume the CGM-data was incorrectly labeled

since the labeling was contradicted by the two other sets.

In the CGM-data, some of the observations had labeling corresponding to a fourth period, except for the three already specified, *lead-in*, *treatment* and *follow-up*. The DEPICT-studies only mentions the three periods and looking at the data the total fraction of observations within this fourth period constituted 1.35 ‰ and without any further information, these observations were discarded since identifying their period and validity was deemed too work intensive.

All recordings in the CGM-data should be on strict five minute intervals. This assumption is the basis for some specific elements of the theory and had to be checked and corrected for. When this was analysed, it was fairly accurate, and 97 % were within the 4-6 minute interval. If this was not corrected for, this still meant that 3 % of all CGM readings had to be discarded. Looking at the gaps in the data, it can be seen that a majority are small, missing only one or two readings, see Figure 3.2.

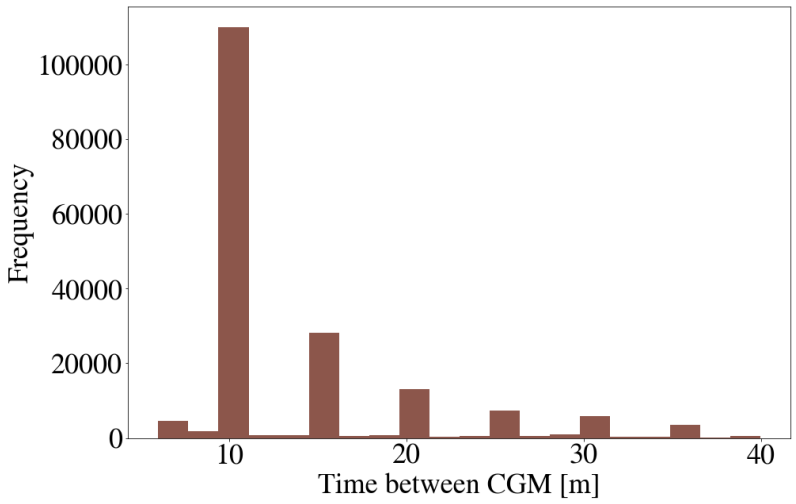


Figure 3.2: Stepwise time difference frequency table for differences larger than six minutes and less than 40 minutes between readings.

Gaps in time series or longitudinal data can be filled using multiple interpolation methods. When the gaps are minor, and the data-frequency surrounding the missing observation is high, linear interpolation can often be sufficient [30]. Up to three values, thus covering gaps of up to 20 minutes, were interpolated to fill gaps in the data. Of the 3.0 percentage points of missing observations, 0.99 percentage points were within the 6 to 22 minutes region. A large majority of the remaining 2.01 percentage points were instead in the 0 to 4 minute region, see Figure 3.3.

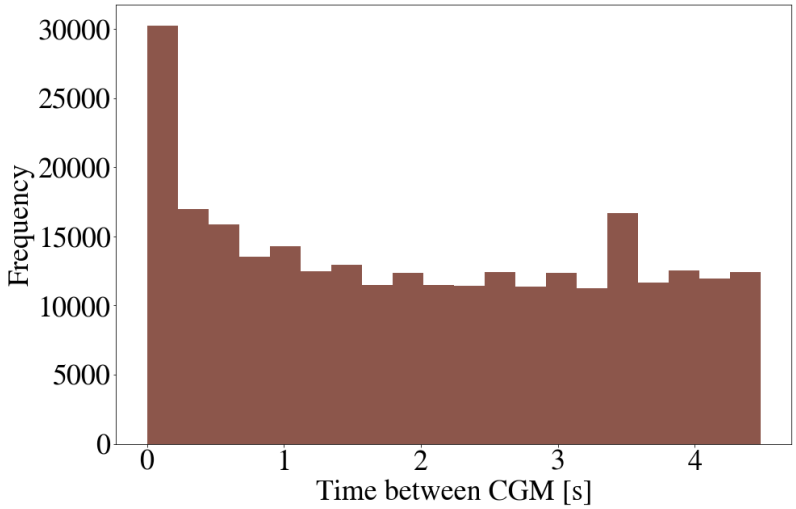


Figure 3.3: Stepwise time difference frequency table for differences less than four and a half minutes between readings.

As can be seen in the figure above, some readings have not registered any time difference. In all these cases, it was deemed impossible to efficiently identify which glucose reading was correct, if they were not of the same value. Because of this, the duplicated readings were removed arbitrarily and accounted for 1.9 ‰ of all observations. The rest of the observations were left untouched since they were not duplicates, and even though not adhering to the strict 5 minute interval, were in the context few.

Some patients, even though labeled as patients who finished the trial, had notably few observations within one or more periods. Since periods should be two weeks with readings every five minutes, having just portions of those approximately 4000 readings over a period could mean these were outliers due to errors. To remove periods with relatively few data points, a cutoff was set at patient's periods with less than a total of three days of data. Threshold was set arbitrarily to minimise the risk of including corrupt data while not unnecessarily excluding too much data. Data within the treatment period is divided into two sub-periods of 2 weeks each, week 11-12 and 23-24. Since both sub-periods of readings within the treatment period are labeled with one label in the raw data, clustering had to be done before the filtering was possible. Clustering was done on the dates of collection of data, which is one dimensional. For the assignment Jenks Natural Breaks algorithm was considered, as described in Section 2.5. Applying Jenks and then setting a threshold of three days excluded 2.0 ‰ of the initial data set.

Within periods there were still outliers which had either wrongly set timestamps or just corrupt data. This was handled by setting a threshold on how long the time difference should be between a cluster of values and another to be a new cluster. Then if any cluster was of less than a certain amount of observations then this cluster was removed. Setting the time difference to 10 minutes and cluster size to 12 excluded 3.8 ‰ of the original data.

In order to capture the dietary habits of the patients enrolled in the study, patients were asked to record diary data of food consumption. More specifically, patients were tasked with recording at which time patients ate breakfast, lunch and dinner. The type of food consumed, calorie intake, or other aspects were not captured. For the dates during which CGM-data were captured, meal-times were only partially recorded. During treatment, CGM-data was recorded for two weeks beginning week 11 and for another two weeks, beginning week 23. Because of this, the total number of days of recorded mealtimes should be 28[1]. This was not the case and patients only registered an average of 12.5 days with mealtimes, see Figure 3.4. Patients were asked to only record meal times for a total of two weeks, during week 12 and week 24.

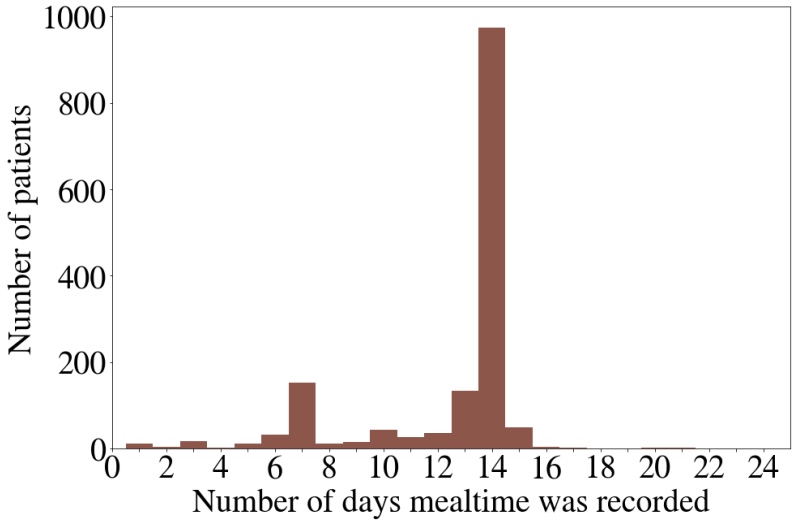


Figure 3.4: Histogram on how many days the patients recorded one or more mealtimes.

A total of 1523 patients recorded at least one mealtime, where the patients recorded mealtimes for 12.5 days on average. The patients could not record if they chose not to eat for the entirety of a day, so it is not possible to distinguish between patients who chose to not eat for a full day, or did not record mealtimes for a full day. It is assumed that the latter is incredibly more common. Moreover, if a patient recorded one or two of the day's three meals, it is assumed that the patient did not eat the meals which were not recorded for that day. During the days where at least one mealtime was recorded, the patients recorded an average of 2.89 mealtimes.

CGM-data was recorded for 28 days during treatment, but mealtimes only for 12.5 days, on average. In order to not be forced to discard over half the data, in the case that *time since last meal* was to be used in the modelling, imputation or removal of missing values had to be done. The nature of mealtimes on a patient level basis is practical in the sense of imputation, since habits and daily routines are fairly consistent. If roughly two weeks of mealtimes are recorded per patient, it is assumed that

[1] $2 \text{ periods} \cdot 2 \text{ weeks} \cdot 7 \text{ days} = 28$

the remaining two weeks can be imputed by calculating the daily average breakfast, lunch and dinner times for each patient. Moreover, since the total average mealtime differs heavily depending on weekday, this has to be taken into account as well. Based on the above, the missing days of recorded mealtimes were imputed as follows. For each patient, the average breakfast, lunch and dinner time for all seven weekdays were computed and saved into a table. If a patient did not have any recordings for a specific meal on a specific weekday, the population average over all patients for that specific meal and weekday was used instead. Because of this, a table containing, for each patient and for all seven weekdays, an average mealtime for breakfast, lunch and dinner existed. For all days where recordings of mealtimes are missing, this table was queried and the patient-specific averages were used in order to impute the missing mealtimes.

During the periods where the patients recorded CGM-data, they were asked to record their daily insulin dosage consisting of both their basal and bolus doses. In the data, it was noticed that patients in rare occasions reported their basal and bolus dosages more than once per day. It is assumed that this was a mistake, and the average of their daily measurements were used as the daily value. Since these dosages are self-reported from patients, some discrepancies in the data quality might exist. Since the analyses of both basal and bolus doses are of interest, there is a requirement for the observations to contain both the basal and bolus doses. In total, of all the days where CGM-data was recorded for all patients, only 88.51 % of the days had both basal and bolus recordings. Thus 11.49 % of the observations were removed in order to allow for the modelling of basal and bolus.

There existed patients in the treatment period who had not registered any data in the lead-in period. Since no baseline for these patients was possible to get these were excluded from the analysis, a total of 40 of the 1480 patients were excluded.

3.2.2 Transformation

Proposed approach to model using logistic regression to predict state changes means that a data set with this feature together with all other features has to be generated. The transformation phase is divided into two separate steps. First merging of all data and then transformation to a data set that can be used by chosen modelling approach. Merging of the four data sets was done on two covariates, the *patient identifier* together with *date* and *time*. Using a Markov model, the goal is to, given a current state, predict the probability of transitioning to any possible state. Enabling this type of modelling on the data, a new target column was generated upon merging. When merging each set of neighbouring CGM readings of size n is reduced into a set of CGM transitions of size $n - 1$, see Table 3.3 and 3.4.

Table 3.3: Pre-transformation table.

ID	Time	...	Present state
1	...:10:15	...	1
1	...:10:20	...	1
1	...:10:25	...	2

Table 3.4: Post-transformation table which has one less row.

ID	Time	...	State change
~~1~~	~~...:10:15~~	...	-
1	...:10:20	...	s_{11}
1	...:10:25	...	s_{12}

At generation of the new feature, as described above, it could happen that the patient got state changes from one state to a non-neighbouring state. Inspecting the data where transition times are within the range 4 to 6 minutes the variance in the changes in glucose readings is 36, see Figure 3.5. As long as the number of states are relatively few it is unlikely that transitions in states will skip states given presented data. By removing observations, in a three state setting, where transitions are to non-neighbouring states data accounting for 0.07 ‰ of the original data was removed. Increased number of states means large increases in the removed observations and should be checked thoroughly when increasing the number of states or creating states in narrow ranges.

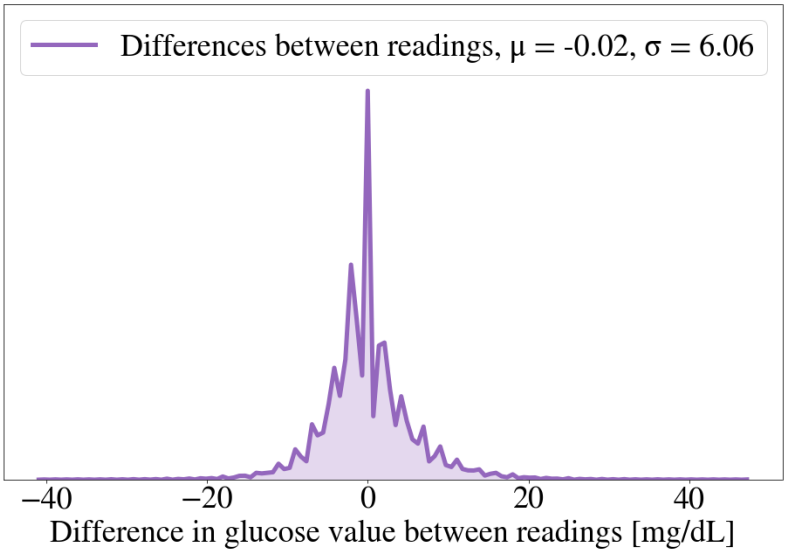


Figure 3.5: Probability density plot for difference between readings in the interval 4 to 6 minutes within the interval from the 0.1 th to the 99.9 th percentile, mean and standard deviation is for the full data.

3.2.3 Feature Extraction

Feature extraction is the process of deriving covariates from other present ones which brings more value to the model. This is done under the hypothesis that these derived ones are are non redundant which means logically explainable ones are the ones generated.

Insulin is, with its direct effect on glucose, a feature of extra significance in this study. The need of insulin depends on several factors as mentioned earlier, some

of which are not in the data, such as physical exercise. Data was though collected for each individual prior to the introduction of the treatment, thus insulin usage in treatment can be normalised using the pre-treatment, baseline, insulin usage. Correlation between the two is good, see Figure 3.6 and 3.7. This covariate can both be introduced as a feature in models to increase model fit when doing analysis, but also set to variable values as input to generated models to analyse whether there exists an optimal reduction.

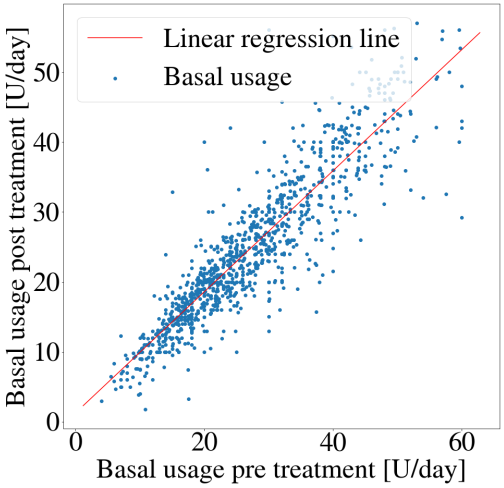


Figure 3.6: Pre-treatment basal insulin to post treatment basal insulin, Pearson correlation coefficient of 0.89. Plotted values for basal are to the 95 th percentile, Pearson correlation coefficient is calculated using all data.

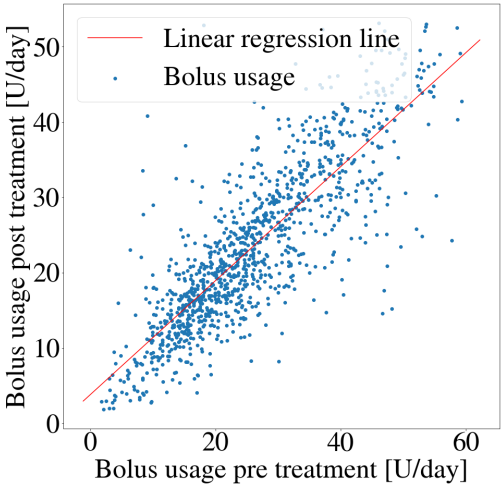


Figure 3.7: Pre-treatment bolus insulin to post treatment bolus insulin, Pearson correlation coefficient of 0.88. Plotted values for bolus are to the 95 th percentile, Pearson correlation coefficient is calculated using all data.

Consumption of food is a crucial aspect of glycemic control, which has been emphasised previously. Since there exists self-reported mealtime data there could be a way to capture this crucial feature. One such way is, for each observation in the CGM-data, to compute the time since last meal. This feature, *time since last meal*, is extracted by looking for the latest meal consumed, given a timestamp. This could in turn be incorporated into a numerical feature, corresponding to the number of minutes since last meal. The way this is computed is shown as pseudo code in Appendix A.

Through explanatory analysis, it could be seen that glucose levels behaved similarly over the days, inhabiting a cyclic behavior. Because of this, the feature *time of day* was extracted, as a categorical feature with 24 levels, one for each hour of the day. This was done by extracting the hour element of the timestamp for each observation.

3.3 Analysis

The multi-state Markov model's main attribute is its transition probability matrix, as described in Equation 2.3. Since the transition probabilities depend on covariates, multiple underlying models have to be created in order to compute the individual probabilities. In this section relating the modelling aspects of this project, "Markov model" relates to the overall model consisting of multiple underlying models.

3.3.1 Selection of Data

In order to be able to fit the Markov model to data, a data set consisting of features and target is required. The data engineering described in Section 3.2 above resulted in a clean and structured data set in a tabular format. This full dataset, consisting of multiple features, is used in the modelling stage. A subset of features would be chosen, based on selection criteria from the user. Below, the different aspects of the selection criteria that exists will be outlined, together with their corresponding effects on the modelling.

When choosing features to be included in the Markov model, both numerical and categorical, actions have to be taken in order for the modelling to be feasible. For the numerical features chosen, normalisation is required in order for the modelling to work properly, as described in Section 2.4.1. Furthermore, the categorical features have to be encoded into dummy-variables, as described in Section 2.4.2. If there is a demand to also model interaction effects within the Markov model, interaction terms between relevant features have to be created and included, as described in Section 2.4.3. When the above is performed, a dataset consisting of the appropriate features is created, and forms the basis for the Markov modelling.

3.3.2 Fit the Markov Model

With the transition probability matrix defined as in Equation 2.3, the probabilities depend on covariates, $\boldsymbol{x}$. In previous literature, this dependency is done with a logistic regression model [15]. Incorporating this into the multi-state Markov model, it is possible to fit a model that predicts a specific state change. Since the probability of transitioning to another neighboring state, or staying in the current one, sums to one, a single logistic regression model has to be fitted for each state. Each model for each state i, depends on a different set of coefficients, $\boldsymbol{\beta}_i$. For the first and last state, where $i = 1$ and $i = n$, there are only two possible transitions as seen in Equation 2.3. This leads to a binomial logistic regression, defined in Equation 2.7, where the probabilities are determined by,

$$p_{\beta_i}(\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\beta}_i^T \boldsymbol{x}}} \qquad i \in \{1, n\}. \tag{3.1}$$

For the other cases, when the current state is not the first nor the last, there are three possible transitions. This leads to a multinomial logistic regression, defined in

Equation 2.12, where the probabilities are determined by,

$$p_i(\boldsymbol{x}) = \frac{e^{\boldsymbol{\beta}_i \boldsymbol{x}}}{\sum_{k=1}^{K} e^{\boldsymbol{\beta}_k \boldsymbol{x}}} \qquad i \in \mathcal{Z} : i \in [2, ..., n-1] \tag{3.2}$$

The n different models are fitted to the data using the SKLearn-module in Python. This module optimises the coefficients $\boldsymbol{\beta}_i$ with the L-BFGS algorithm described in Section 2.3. This way the Markov model is fitted to the data, and its transition probability matrix, $\boldsymbol{P}$, is created, which is dependent on the features included in the model. A representation of the model, when applying it on the given problem with three states, can be seen in Figure 3.8. The method described is applicable to any number of states.

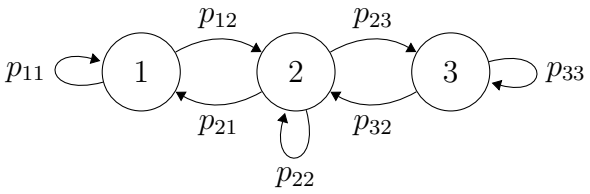


Figure 3.8: Abstract representation of a multi-state Markov model with three states.

3.3.3 Confidence Intervals

In order to get confidence intervals for the MST and fraction of time spent in the different states, bootstrapping was performed. This was done as described in Section 2.6.3, where the number of bootstraps were set to 1000. Line-plots with 95 % confidence levels were generated in order to display the range of confidence.

3.3.4 Validation

The validation was performed utilising Brier score, as explained in theory, Section 2.6.1. The Markov model created consists of multiple logistic regression models depending on the current state, as described in Section 3.3. For clarification, each possible state corresponds to a binomial or multinomial logistic regression model predicting the probability of transitioning to neighboring states. Because of the multitude of underlying models as well as a different number of observations in each state, the computation of the Brier Score has to be adapted in a way that accurately reflects the performance of the Markov model. As described in detail in theory, Section 2.6.1, Brier score is an evaluation of the mean square error between the predicted probabilities assigned to the possible outcomes and the actual outcomes. The range of probabilities, $[0, 1]$, is traditionally divided into a set of equally sized bins, where the comparison is made and each bin is weighted equally. For the use case of this multi-state Markov model on the other hand, where the number of observations are highly concentrated within a tiny range of probabilities, the bins are instead formed based on the number of observations. This results in bins varying in range, but of equal weight in terms of number of observations which gives a

more fair measurement of model performance. A Brier score is computed for each predicted state change, which are then averaged to a single scalar, thus representing the performance of the Markov model.

In order to reliably validate the model, K-fold cross validation was performed as described in Section 2.6.2. The number of folds, K, was chosen to 10, as it has been proven to be successful in previous work [31]. The evaluation metric, Brier score, was computed as described above, once per fold, then averaged over the 10 folds, thus acting as the validation score for the Markov model. This validation score could then be compared between Markov models.

4
Results

Given the previous methodology, results regarding three major areas are presented. First, the adaptation and implementation of the Markov model is described, involving the library specifically created for the task of analysing the DEPICT-data. Secondly, features of interest and their relationship to glycemic control are presented. Lastly, insulin reduction, both basal and bolus, is presented and their relationship to glycemic control is presented.

4.1 Adaptation and Implementation of the Markov Model

Given the high frequency data points from the CGM-data, with 5-minute intervals, it can be shown that all state changes are observed, as motivated in Section 3.2.2. Because of this, only state changes to neighboring states are allowed, which is an adaptation to the original Markov model. Moreover, since there is no progression in some particular direction, there exists no absorbing states. The transition probability matrix is computed around a logistic regression framework, using the L-BFGS algorithm for parameter estimation. This transition probability matrix is then used to compute fraction of time spent in the various states as well as the mean sojourn time, which are used as a basis for analysis.

In order to implement a Markov model the way as described previously in Chapter 3, there was a need to implement a library that given a data set, processed the data accordingly as well as generated and fitted a Markov model to the data. This has been implemented in a library, written in Python. The structure of the code is here outlined in order to give insights into the workings of generating a Markov model.

1. **Load data**
   The data files containing all information gathered during the DEPICT studies are included. More specifically, this involves all data regarding CGM-readings, insulin usage, patient information as well as mealtime data. All these files are loaded into memory.

2. **Choose number of states to model**
   The user specifies how many states are desired to model, and their corresponding glucose limits in mg/dL.

3. **Pre-processing**
   Pre-processing is performed, as described thoroughly in Section 3.2, where data points are removed and added as well as merged together from multiple data sources. Moreover, feature extraction is performed to extract features involving *time since last meal*, *insulin reduction* and *time of day* to name a few.

4. **Choose what time-frame to model on**
   The user then specifies over which time frame the data should be contained. The feature *days on treatment* can be used to specify if the data to be used is concerning the lead-in period, the treatment period, or some other time period.

5. **Choose features to include**
   The user specifies which categorical features and numerical features should be included in the model, where any number of features can be specified.

6. **Add dummy variables**
   Dummy variables were derived using one hot encoding for the categorical features, where each level of each feature becomes a new, binary, feature.

7. **Normalise the data**
   The data is normalised, with a scaler of choice from the user: *minmax, maxabs* or *standard scaler.* The scalar parameters are saved.

8. **Add interaction terms**
   The user can choose to include interaction terms, which are then generated by a process of multiplication, resulting in additional features.

9. **Initiate the Markov model**
   Finally, a Markov model with the specifics from the previous steps is initiated and fitted to the data.

When the Markov model is initiated and fitted to the data, as described above, it can be utilised for analysis. Below the different aspects of analysis are described.

- **Perform model validation**
  Use the Markov model and perform K-fold cross validation as described in Section 2.6.2. This gives the Brier score, thus the evaluation metric, for the model and can be compared with the Brier score from Markov models based on a different set of features.

- **Create patient groups**
  In order to compare different patient groups and gain insights, patient groups with different values for different features can be created. If, for example, comparison of the different treatment arms was of interest, three different patient groups would be created.

- **Transition probability matrix**
  The transition probability matrix, $\mathbf{P}$, for each patient group is computed, which consists of the probabilities of transitioning between the different states. This matrix is then used for further analysis.
- **Fraction over day and mean sojourn times**
  The transition probability matrix, $\mathbf{P}$, computed above can be used to compute the fraction over day as well as mean sojourn times for each patient group, as described in Section 2.2.2 and 2.2.1. This way, the two metrics for the different patient groups can be compared and analysed, which enables well-founded conclusions to be drawn.

4.1.1 Feature Selection

During the modelling phase, multiple models are used to draw insights into various topics. In a general sense, a model should have as high fit as possible which often results in adding many features. However, when a specific relation between one or many features and the outcome is analysed, only this specific set of features is needed in the model. This because, for the specific analysis in question, all other contributions to the outcome is not of interest and are thus included as noise in the model.

The kidney function of patients is highly relevant for the effect of dapagliflozin. Because of this, the relationship between Glomerular Filtration Rate (GFR), which is an indicator of kidney function, and glycemic control is chosen to be analysed. Another interesting aspect is the time on treatment, which might have a relationship to the glycemic control of the different treatment arms. Because of this, time on treatment is also chosen to be analysed.

Two features, *time of day* and *time since last meal*, were extracted which were supposed to capture similar behavior in the data, periodicity during the day due to eating. Time of the day might capture other periodic aspects of an individual's life, it might also quite badly fit the meals which will not be fixed to a certain time each day. Mealtime on the other hand might, because of the already quite poor data quality, not capture the effect. To analyse which of these is the better feature, Brier score was used. A simple model with only *treatment arm* and one of the two other features was fitted which yielded the Brier scores in Table 4.1. The Brier score was calculated using ten-fold cross validation.

Table 4.1: Comparison of fit between two simple models with the only difference of one feature, either *time of day* or *time since last meal*.

Model Number	Feature	Brier score
1	*Time of day*	$352.2 \cdot 10^{-4}$
2	*Time since last meal*	$352.7 \cdot 10^{-4}$

Looking at the Brier score for the two models one can see that model 1 is closer to the perfectly calibrated model. The feature *time of day* fit better to the data.

4.2 General Features' Relation to Glycemic Control

When features that might have a relation to glycemic control, other than dapagliflozin and insulin, were explored, the patients' kidney function was particularly interesting. The relationship between GFR and glycemic control was modelled in order to enable analysis. The relationship between fraction of stay in each state, dependent on kidney function, measured in GFR, could show if there are any dependencies on the well-being of patients on the treatment arms dependent on their kidney function. A Markov model is fitted with the interaction between GFR and *treatment arm* as a feature. The analysis is then made by comparing the length of stay, compared with baseline, for each treatment arm, depending on GFR value. Given this analysis, no apparent trends can be seen, see Figure 4.1. The baseline that the fraction of stay in each state is compared against, is computed by splitting the pre-study data dependent on thresholds in GFR.

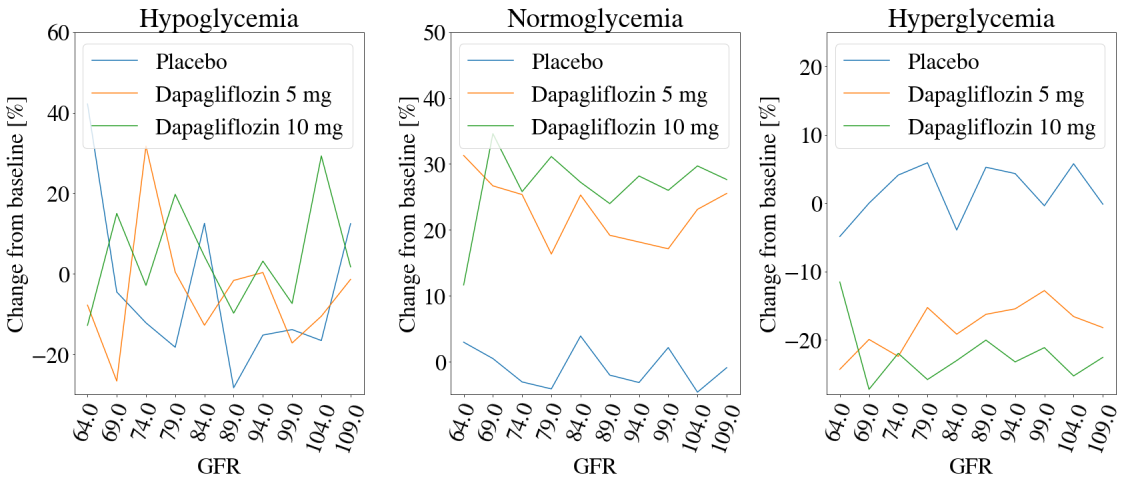


Figure 4.1: Stay in each glycemia state compared to baseline for different levels of GFR values. The result is generated by fitting the Markov model using the interaction between GFR and *treatment arm* as a feature over a given GFR range.

4.3 Reduction of Insulin

Patients in both treatment arms, dapagliflozin 5 mg and dapagliflozin 10 mg, reduced their usage of insulin compared to baseline, see Figure 4.2 and 4.3. During the DEPICT-studies, patients were recommended to reduce their insulin by up to 20 % at the start of the treatment period and then urged to resume medication to previously established individual baseline levels. Analysing the time spent in each state as well as mean sojourn times for the different states becomes highly interesting to see if any specific reduction corresponded to better reaction to dapagliflozin. The span of reduction to be analysed is set to vary symmetrical among basal and bolus, between 0 % and 24 %, for which there exist a substantial amount of data. Thus the model is not extrapolating out of data ranges. Worth noting is that patients had a

large amount of days where they chose not to decrease their basal insulin, looking at the peak at 1.0 in Figure 4.2. A large portion of patients in the treatment arm of 10 mg of dapagliflozin had days where they reduced their basal insulin by exactly 20 % compared to baseline. There is no data in these figures from the first few days of treatment, only from week 11, 12, 23 and 24, thus the initial recommended reduction would not affect the distribution plots.

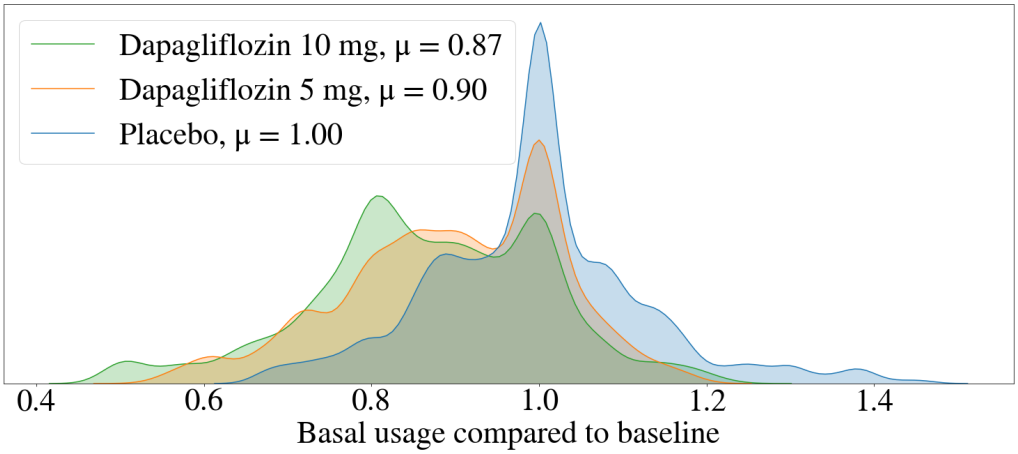


Figure 4.2: Probability density function of daily basal insulin usage compared to baseline in the treatment periods for each treatment arm. This is with data for each treatment arm within the interval from the 2.5th to the 97.5th percentile, mean values are for the full data.

Reduction in bolus insulin was more normally distributed, where patients in both treatment arms reduced their usage by 10 % and 8 % on average, respectively, while the placebo group instead increased usage by 3 % compared to baseline, see Figure 4.3.

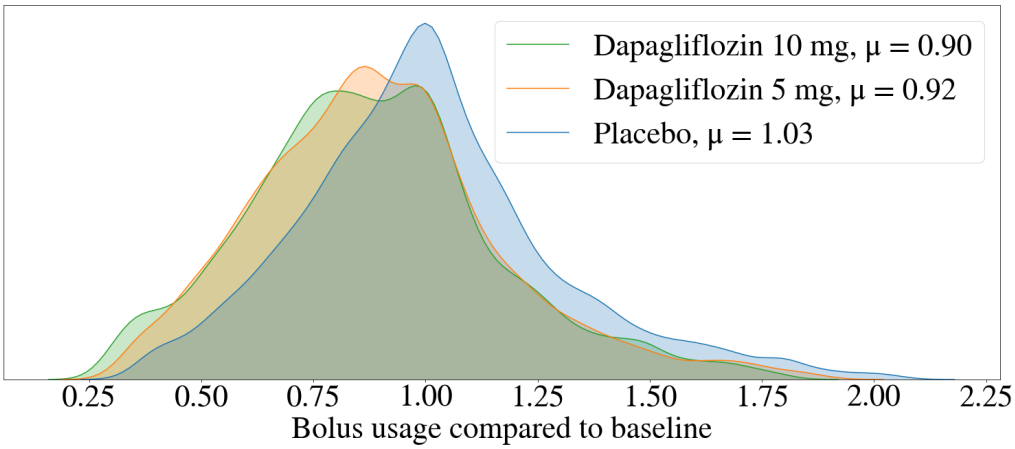


Figure 4.3: Probability density function of daily bolus insulin usage compared to baseline in the treatment periods for each treatment arm. This is with data for each treatment arm within the interval from the 2.5th to the 97.5th percentile, mean values are for the full data.

Modelling the effect of reducing insulin on each treatment arm shows patients on the two treatment arms, dapagliflozin 5 mg and dapagliflozin 10 mg, increase time spent in normoglycemia compared to baseline, see Figure 4.4. Moreover, the model shows that time spent in hyperglycemia reduces for the two treatment arms involving dapagliflozin as well. It should be noted that the portion of day spent in the hypoglycemic state consists of very low fractions as a baseline. Because of this, the reduction of 10 % as seen for placebo and dapagliflozin 5 mg corresponds to a reduction, from baseline, of roughly 0.5 percentage points in hypoglycemia.

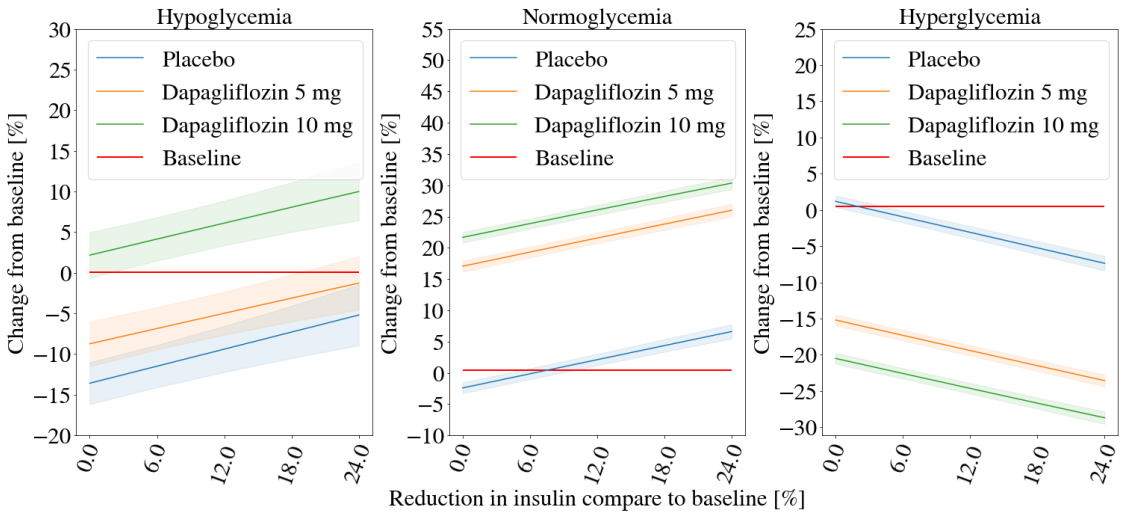


Figure 4.4: Fraction of day with a 95 % confidence interval compared to baseline, based on treatment arm and reduction in insulin.

4.3.1 Fraction of Day Spent in Each Glycemic State

Line-plots with 95 % confidence intervals, the shaded area, are formed to show the fraction of day spent in the different glycemic states, and to give an estimation of the level of confidence, given different levels of insulin reduction. The model shows that there is an increased amount of time spent in normoglycemia when insulin is reduced, see Figure 4.5. A similar relation to insulin reduction can also be seen for time in hyperglycemia, where reducing insulin reduces the time spent in that state. The effect is less substantial in the case of hypoglycemia, where the effect is less beneficial for the patient, since time in this state is increased when reducing insulin. It should be noted however, that the number of percentage points difference is far from as high, in comparison with the benefits in normo- and hyperglycemia.

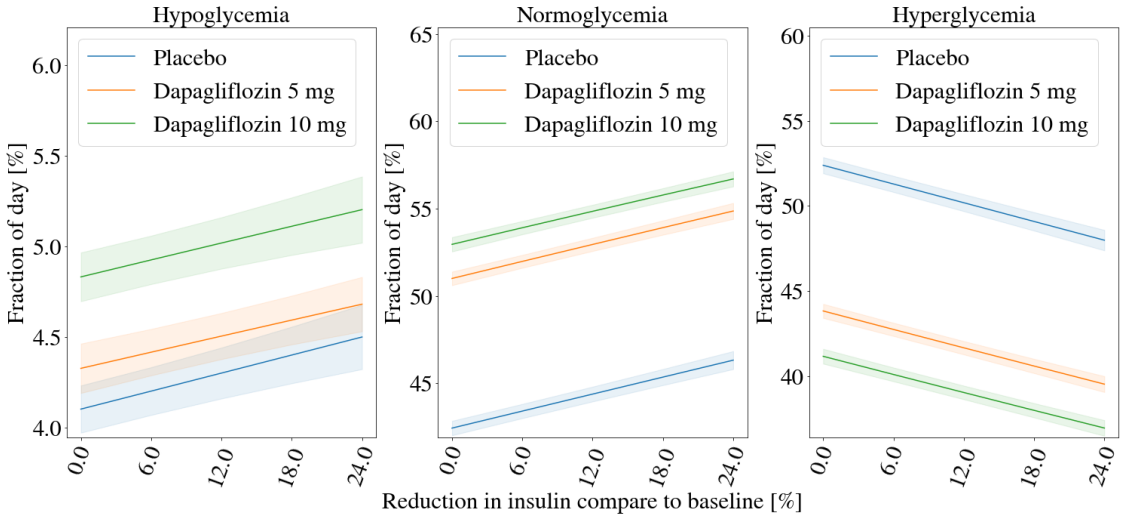


Figure 4.5: Fraction of day with a 95 % confidence interval spent in different states, based on treatment arm and reduction in insulin.

4.3.2 Mean Sojourn Times

The mean sojourn time in each state is the expected amount of time spent there once entered. Line-plots with 95 % confidence intervals, the shaded area, are formed to show the mean sojourn times in the different glycemic states, and to give an estimation of the level of confidence, given different levels of insulin reduction. The relationship between mean sojourn time and insulin reduction is of the same character as for the fraction of day in the previous section, Section 4.3.1. However, the sojourn time in minutes for the different states, dependent on insulin reduction and treatment arm, can be seen below in Figure 4.6

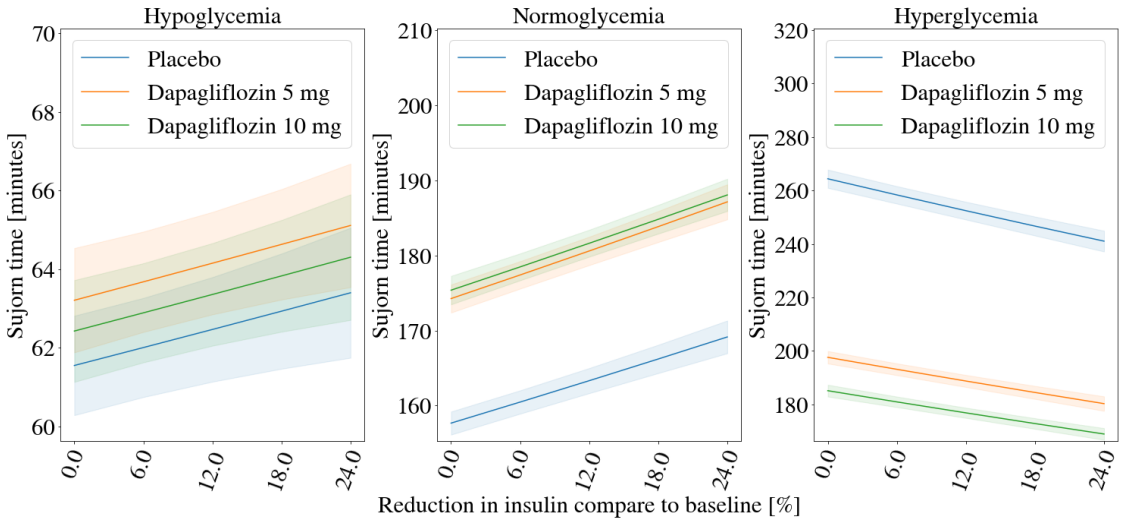


Figure 4.6: Mean sojourn time with a 95 % confidence interval spent in different states, based on treatment arm and reduction in insulin.

4.3.3 Five-State Model

The previous three-state model suggests that when insulin is reduced, the time in hyperglycemia is reduced, as described in Section 4.3.1. In order to explore if further insights could be had within this hyperglycemic region, a more granular Markov model with more states was created. The previous glucose limit for a hyperglycemic state was set as >180 mg/dL. The glucose limits for the five states were now instead set to '<70 mg/dL', '70-180 mg/dL', '180-220 mg/dL', '220-260 mg/dL' and '>260 mg/dL'. The model now consisted of five states, where the granularity in the upper region of glucose measurements was increased, allowing for inference in the extreme hyperglycemic range to be had. This five-state model suggests that the relation with insulin reduction in the highest glycemic state, '>260 mg/dL', has an even greater correlation on time spent in this extreme hyperglycemic state, compared to its lower states, '180-220 mg/dL' and '220-260 mg/dL' respectively, see Figure 4.7. This even greater correlation on time spent can be seen by noting that the angle of the slope of the lines are larger in the highest glycemic state, '>260 mg/dL', compared with the glycemic states '180-220 mg/dL' and '220-260 mg/dL'.

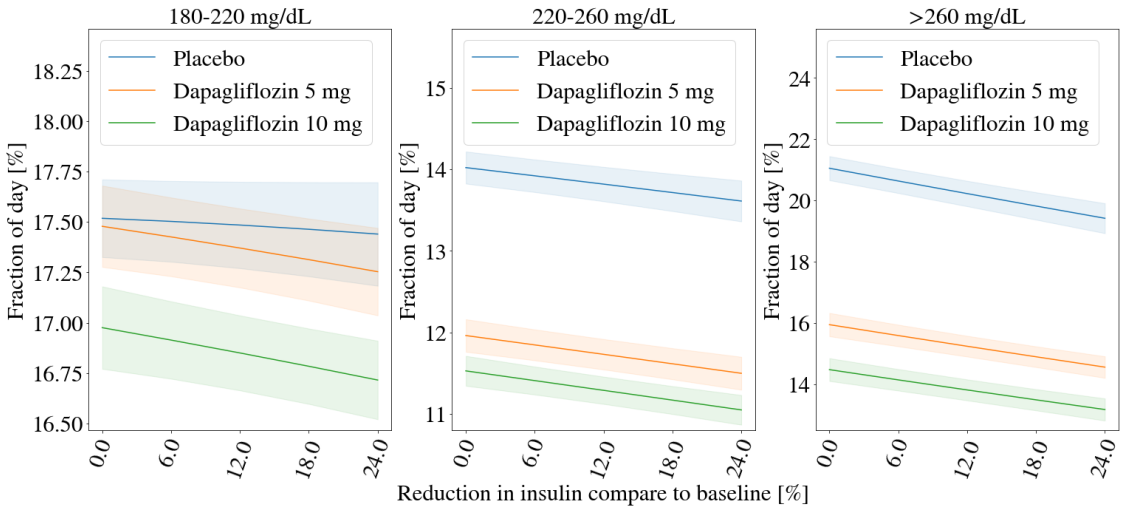


Figure 4.7: Fraction of day with a 95 % confidence interval spent in three extreme-hyperglycemic states '180-220 mg/dL', '220-260 mg/dL' and '>260 mg/dL', based on treatment arm and reduction in insulin.

4.3.4 Other Model Variations

The Markov model was, in addition to the reduction of insulin, also fitted with an interaction term between bolus and basal reduction, thus allowing for more complex interaction effects to be captured. This model did however not yield any other insights than the ones presented previously, in Section 4.3.1. Moreover, another approach tried was to not set the reduction of bolus and basal symmetric. Since the patients did not change their basal dosages for a a large majority of days, as seen in Figure 4.2, this reduction was set to 0, while the bolus was reduced between 0 % and 24 %. However, neither this did result in any insights other than the ones presented previously in Section 4.3.1.

5
Discussion and Conclusions

Given the data collected from the DEPICT-studies, which have served as a basis for this project, the multi-state Markov model has proven possible to adapt in a way that incorporates the different aspects of interest. Developing the model around a logistic regression framework, using the L-BFGS algorithm for parameter estimation, have shown to be successful in modelling a set of defined states. The assumption that all state changes are observed, given the high frequency data points, has proven to be applicable when estimating the transition probability matrix. Based on this assumption, fraction of day spent in each state as well as mean sojourn times were successfully computed. The library developed around the Markov model allows for flexibility in regards to the number of states chosen, what features to be included as well as the computation of performance metrics to be validated against. This performance metric, using Brier scores, was proven useful although challenging to implement successfully since it is mainly used for feature selection and is hard to use in order to determine how well a model fits the underlying data. There could be other, even more useful, measurements of model performance as well as model fit, applicable for this type of Markov models. The Markov model developed during this project and its applications have thus proven to be a useful and flexible tool in the inference of features affecting glycemic control.

As previously established in studies, patients receiving dapagliflozin are experiencing significant improvements in glycemic control. The modelling shows that patients on dapagliflozin experience large increases in the portion of time spent in normoglycemia, large decreases in the portion of time spent in hyperglycemia while experiencing only slight increases in portion of time spent in hypoglycemia, in comparison with placebo patients. Capturing the effect of fluctuations in glucose over the day due to eating was better approximated using *time of day* than *time since last meal*. *Time since last meal* could be used complementary to the *time of day* feature since it is possible that the time of the day captures other periodically recurring activities, such as physical activity. Modelling effects using the time since the actions was successful though and gives an indication on how to successfully integrate the effect of eating, or other activities, in future models. Measured GFR does not seem to have any relation to how well the different levels of medication work for patients with normal kidney function, GFR values within 60 to 110. Given that the patients were exposed to dapagliflozin during relatively long time frames, the time spent in the different states could change over time.

There is a causal relationship between insulin and decreased glucose levels. Thus patients, or days where patients, used more insulin than usual should mean a decreased portion of time spent in hyperglycemia and an increased time in hypoglycemia compared to days of using less insulin. However, this is not what was found modelling the effect, for any of the treatment arms, dapagliflozin nor placebo. Instead, with a decrease in insulin usage one can see decreased time in hyperglycemia and a slight increase in hypoglycemia, opposite of what could be assumed. It should be noted that patients on the two treatment arms receiving dapagliflozin experienced better glycemic control over the entire range of insulin reduction that is modeled, compared to the placebo arm. Moreover, a five-state model was created to get more insights into the extreme-hyperglycemic range, '>260 mg/dL'. The model shows that in this extreme region, the relationship between the reduction of insulin and decreased time in extreme hyperglycemia was even stronger. The causal relationship is well established and something must instead be masking the effect. The reason for this could be because of the lack of granularity within the data regarding insulin doses, food consumption and physical exercise, as mentioned in the limitations of this project. One explanation could thus be that this relationship discovered is due to the fact that insulin is only a proxy for eating unevenly; a day of increased insulin would correspond to a day with worsened glycemic control since this was a day of abnormally large food and drink intake.

5.1 Future Work

The chosen performance metric, Brier score, does not perform well given how it is applied in this project. It is mainly used for feature selection, and not for how well models fit to the underlying data, since this was notoriously hard to implement. However, this project mainly focused on the relationship between specific features and glycemic states, which resulted in that comparison of different models based on a selection of varying features was never needed. Implementation of other more appropriate performance metrics allowing evaluation of model fit would bring great value.

This project has focused on modelling glucose levels based on various data that is thought to have a glycemic effect. However, it is previously known that two major drivers of glucose changes, in diabetes patients, are food consumption and insulin intake. Information about how much, and what kind, of food is consumed together with exact times for administration of insulin and the corresponding dosage is thus crucial for glycemia control. Possibly even data on other activity with relations to glycemic control, like physical activity, would be needed to make the attempted analysis plausible. If such increased granularity within the data could be achieved in the future, more nuanced modelling could take place. With such a model, hopefully an even more accurate relationship between dapagliflozin and reduction of insulin could be established.

Bibliography

[1] S. Lehto, T. Rönnemaa, S. M. Haffher, K. Pyörälä, V. Kallio, and M. Laakso, "Dyslipidemia and hyperglycemia predict coronary heart disease events in middle-aged patients with niddm," Diabetes, vol. 46, no. 8, pp. 1354–1359, 1997.

[2] W. H. Pan, L. B. Cedres, K. Liu, A. Dyer, J. A. Schoenberger, R. B. Shekelle, R. Stamler, D. Smith, P. Collette, and J. Stamler, "Relationship of clinical diabetes and asymptomatic hyperglycemia to risk of coronary heart disease mortality in men and women," American journal of epidemiology, vol. 123, no. 3, pp. 504–516, 1986.

[3] F. H. Epstein, "Special article," Circulation, vol. 36, no. 4, pp. 609–619, 1967.

[4] L. S. Williams, J. Rotich, R. Qi, N. Fineberg, A. Espay, A. Bruno, S. E. Fineberg, and W. R. Tierney, "Effects of admission hyperglycemia on mortality and costs in acute ischemic stroke," Neurology, vol. 59, no. 1, pp. 67–71, 2002.

[5] P. J. Lindsberg and R. O. Roine, "Hyperglycemia in acute stroke," Stroke, vol. 35, no. 2, pp. 363–364, 2004.

[6] B. F. Schrijvers, A. S. De Vriese, and A. Flyvbjerg, "From Hyperglycemia to Diabetic Kidney Disease: The Role of Metabolic, Hemodynamic, Intracellular Factors and Growth Factors/Cytokines," Endocrine Reviews, vol. 25, pp. 971–1010, 12 2004.

[7] R. Lee, T. Y. Wong, and C. Sabanayagam, "Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss," Eye and vision, vol. 2, no. 1, p. 17, 2015.

[8] P. J. Lindsberg and R. O. Roine, "Hyperglycemia in acute stroke," Stroke, vol. 35, p. 363–364, 2004.

[9] K. Torimoto, Y. Okada, H. Mori, and Y. Tanaka, "Relationship between fluctuations in glucose levels measured by continuous glucose monitoring and vascular endothelial dysfunction in type 2 diabetes mellitus," Cardiovascular diabetology, vol. 12, no. 1, p. 1, 2013.

[10] J. W. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the adap learning algorithm to forecast the onset of diabetes mellitus," in Proceedings of the Annual Symposium on Computer Application in Medical Care, p. 261, American Medical Informatics Association, 1988.

[11] C. Mathieu, P. Dandona, P. Gillard, P. Senior, C. Hasslacher, E. Araki, M. Lind, S. C. Bain, S. Jabbour, N. Arya, et al., "Efficacy and safety of dapagliflozin in patients with inadequately controlled type 1 diabetes (the depict-2 study):

24-week results from a randomized controlled trial," Diabetes Care, vol. 41, no. 9, pp. 1938–1946, 2018.

[12] P. Dandona, C. Mathieu, M. Phillip, L. Hansen, S. C. Griffen, D. Tschöpe, F. Thorén, J. Xu, A. M. Langkilde, J. Proietto, et al., "Efficacy and safety of dapagliflozin in patients with inadequately controlled type 1 diabetes (depict-1): 24 week results from a multicentre, double-blind, phase 3, randomised controlled trial," The lancet Diabetes & endocrinology, vol. 5, no. 11, pp. 864–876, 2017.

[13] Food and D. Administration, "Farxiga® (dapagliflozin) tablets, for oral use initial u.s. approval: 2014," 2014.

[14] E. M. Agency, "Forxiga," 2012.

[15] R. Kay, "A markov model for analysing cancer markers and disease states in survival studies," Biometrics, vol. 42, no. 4, pp. 855–865, 1986.

[16] I. M. Longini Jr, W. S. Clark, R. H. Byers, J. W. Ward, W. W. Darrow, G. F. Lemp, and H. W. Hethcote, "Statistical analysis of the stages of hiv infection using a markov model," Statistics in medicine, vol. 8, no. 7, pp. 831–843, 1989.

[17] P. K. Andersen, "Multistate models in survival analysis: a study of nephropathy and mortality in diabetes," Statistics in medicine, vol. 7, no. 6, pp. 661–670, 1988.

[18] R. Gentleman, J. Lawless, J. Lindsey, and P. Yan, "Multi-state markov models for analysing incomplete disease history data with illustrations for hiv disease," Statistics in medicine, vol. 13, no. 8, pp. 805–821, 1994.

[19] S. W. Duffy, H.-H. Chen, L. Tabar, and N. E. Day, "Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and exit from the preclinical detectable phase," Statistics in medicine, vol. 14, no. 14, pp. 1531–1543, 1995.

[20] G. Marshall and R. H. Jones, "Multi-state models and diabetic retinopathy," Statistics in medicine, vol. 14, no. 18, pp. 1975–1983, 1995.

[21] I. W. Marsh, "High-frequency markov switching models in the foreign exchange market," Journal of Forecasting, vol. 19, pp. 123–134, 2000.

[22] D. Cziraky and D. Zink, "Multi-state markov modelling of ifrs9 default probability term structure in ofsaa (industry report)," London, UK: Oracle, 2017.

[23] S. Y. (https://math.stackexchange.com/users/265986/sunghee yun), "Logistic regression - prove that the cost function is convex." Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/3198681 (version: 2019-04-23).

[24] J. N. Dong C. Liu, "On the limited memory bfgs method for large scale optimization," Mathematical Programming, vol. 45, p. 503–528, 1980.

[25] J. H. Klotz and L. D. Sharples, "Estimation for a markov heart transplant model," Journal of the Royal Statistical Society: Series D (The Statistician), vol. 43, no. 3, pp. 431–438, 1994.

[26] T. Duprey and B. Klaus, How to predict financial stress? An assessment of Markov switching models. ECB Working Paper, 2017.

[27] P. Delanaye, E. Schaeffner, N. Ebert, E. Cavalier, C. Mariat, J.-M. Krzesinski, and O. Moranne, "Normal reference values for glomerular filtration rate: what do we really know?," Nephrology Dialysis Transplantation, vol. 27, pp. 2664–2672, 07 2012.

[28] M. Jessica R. Weinstein and M. Sharon Anderson, "The Aging Kidney: Physiological Changes," Advances in Chronic Kidney Disease, vol. 17, no. 4, pp. 302–307, 2010.
[29] Dexcom, "Continuous glucose monitoring system - user's guide," 2015.
[30] F. H. C. Mathieu Lepot, Jean-Baptiste Aubin, "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment," Water, vol. 9, no. 10, p. 796, 2017.
[31] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," Ijcai, vol. 14, no. 2, pp. 1137–1145, 1995.

A

Time Since Last Meal

Algorithm 1: Given CGM-data and meal times, compute time since last meal

Result: Time since last meal for all observations

Define *meal_times* as a table consisting of information regarding if and when each patient have eaten their Breakfast, Lunch and Dinner, as described in section 3.2.3;

for each patient in CGM-data **do**
    Take a subset of the CGM-data corresponding to the specific patient;
    for each day of recordings in the subset **do**
        Take a subset corresponding to the specific date.;
        Define *todays_meals* as the mealtimes from *meal_times* corresponding to the current date;
        Define *yesterdays_meals* as the mealtimes from *meal_times* corresponding to the current date minus one day;
        for each observation **do**
            Define *timestamp* as the date and time when the observation was recorded;
            if *todays_meals*['Dinner'] happend before *timestamp* **and** *todays_meals*['Dinner'] was eaten **then**
                time_since_last_meal ← *timestamp* - *todays_meals*['Dinner']
            else if *todays_meals*['Lunch'] happend before *timestamp* **and** *todays_meals*['Lunch'] was eaten **then**
                time_since_last_meal ← *timestamp* - *todays_meals*['Lunch']
            else if *todays_meals*['Breakfast'] happend before *timestamp* **and** *todays_meals*['Breakfast'] was eaten **then**
                time_since_last_meal ← *timestamp* - *todays_meals*['Breakfast']
            else if *yesterdays_meals*['Dinner'] was eaten **then**
                time_since_last_meal ← *timestamp* - *yesterdays_meals*['Dinner']
            else if *yesterdays_meals*['Lunch'] was eaten **then**
                time_since_last_meal ← *timestamp* - *yesterdays_meals*['Lunch']
            else
                time_since_last_meal ← *timestamp* - *yesterdays_meals*['Breakfast']
            end if
        end
    end
end

B

Jenks Natural Breaks Algorithm

Algorithm 2: Jenks Natural breaks

Result: A set of breakage points in the data
Calculate mean of data;
Calculate the SDAM 2.14;
for each possible split **do**
    for each group in that split **do**
        Calculate group mean;
        Calculate the SDCM;
    end
    Sum all the SDCM to SSDCM 2.15;
end
for each possible split **do**
    Calculate the score: (SDAM — SSDCM) / SDAM;
end
Highest score is the best split.