

The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas

Melanie K. Bothe¹, Luke Dickens², Katrin Reichel^{1,3}, Arn Tellmann¹, Björn Ellger⁴,
Martin Westphal^{1,4}, A. Aldo Faisal^{2,3}

¹Fresenius Kabi Deutschland GmbH, Else-Kröner-Strasse 1, 61352 Bad Homburg, Germany;

²Department of Computing, Huxley Building, Imperial College London, London SW7 2AZ,

UK; ³Department of Bioengineering, Royal School of Mines Building, Imperial College

London, London SW7 2AZ, UK; ⁴University Hospital, Department of Anesthesiology,

Intensive Care Medicine and Pain Therapy, 48149 Muenster, Germany

Running title: Reinforcement learning algorithms in diabetes care

Corresponding authors: Melanie K. Bothe (physiology), Fresenius Kabi Deutschland GmbH, Else-Kröner-Strasse 1, 61352 Bad Homburg, phone: +49 (0)6172-608-4556, fax: +49 (0)6172-686-7291, mail: melanie.bothe@fresenius-kabi.com; **A. Aldo Faisal** (bioengineering), Department of Computing and Department of Bioengineering, Imperial College London, London SW7 2AZ, UK, phone: +44 (0)781-5863146, fax: +44 ... , mail: a.faisal@imperial.ac.uk

Key words: artificial pancreas, continuous glucose monitoring, diabetes mellitus, machine learning, personalized medicine, reinforcement learning

Abbreviations: ANN = artificial neural network, CGM = continuous glucose monitoring, GIP = gastric inhibitory peptide, GLP-1 = glucagon-like peptide 1, HbA_{1c} = glycosylated hemoglobin, ICU = intensive care unit, MPC = model predictive control, PID = proportional integral derivative, RL = reinforcement learning

Summary

Blood glucose control, e.g. in diabetes mellitus or severe illness, requires strict adherence to a protocol of food, insulin-administration and exercise personalized to each patient. An artificial pancreas for automated treatment could boost quality of glucose control and patients' independence. The components required for an artificial pancreas are (1) continuous glucose monitoring (CGM), (2) smart controllers and (3) insulin pumps delivering the optimal amount of insulin. In recent years, medical devices for CGM and insulin administration have undergone rapid progression and are now commercially available. Yet, clinically available devices still require regular patient or care-givers attention as they operate in open-loop control with frequent user intervention. Dosage-calculating algorithms are currently being studied in intensive care patients , for short overnight control to supplement conventional insulin delivery , and for short periods where patients rest and follow a prescribed food regime . Fully-automated algorithms that can respond to the varying activity levels seen in outpatients, with unpredictable and unreported food intake, and which provide the necessary personalized control for individuals is currently beyond the state-of-the-art. Here we review and discuss reinforcement learning algorithms, controlling insulin in a closed-loop to provide individual insulin dosing regimens that are reactive to the immediate needs of the patient.

Introduction

Maintaining normoglycemia is one of the major challenges in the treatment of patients with diabetes mellitus. As a key performance indicator, an average plasma glucose concentration of glycosylated hemoglobin (HbA_{1c}) values < 7% was recommended by the American Diabetes Association and has been shown to reduce the development and progression of microvascular and cardiovascular complications by 76% . Conversely, treatment of hyperglycemia with insulin may lead to hypoglycemia that, in turn, may contribute to clinically relevant complications. This implies that calculation of precise insulin dosages is critical, must be individually adapted, and should be reactive to the patient's glucose level. Worldwide, more than 371 million people have to manage their diabetes with both constant glucose monitoring and insulin dosing that affect their quality of life. This has led to intensive research concerning the development of an artificial pancreas since the 1970s . Such a system is composed of three components: 1. continuous glucose monitoring (CGM) using an implanted sensor, 2. an insulin pump delivering insulin, and 3. an algorithm calculating the correct dose of insulin to be applied. Since the development of the first artificial pancreas system, major improvements have been made, but the system still needs development before it can be routinely used in clinical practice. To date, one of the major limitations of the successful use of automated dosing in clinical, as well as outpatient settings, is the demand for a flexible algorithm that adapts the artificial pancreas to the special needs of each single patient. In this article, we briefly describe the major challenges for development of an artificial pancreas system and discuss the application of machine learning algorithms as a potential approach to increase the flexibility of the system.

Control strategies for insulin delivery

In standard diabetes treatment, the patient receives a subcutaneous injection of slow-acting insulin to provide the basal insulin requirement. Additional insulin doses of rapid-acting insulin are calculated based on the patient's knowledge of a meal size, the patient's experience, the insulin sensitivity and actual blood glucose levels measured indirectly, for example by a subcutaneous CGM system. Insulin boluses are preferably delivered by an insulin pump to avoid repeated injections. In the literature, this model of calculating an insulin dosage based on blood glucose levels and external (meal) information has been called open-loop control . Despite the advantages of the intended avoidance of hyper- and hypoglycemic excursions by use of an artificial pancreas system, any open-loop insulin control mechanism requires the patient to live a more or less predictable lifestyle. In contrast, closed-loop control models include the benefit to finally reduce the patients' need to plan each day with regard to their illness, thereby improving quality of life (for review about open- and closed-loop models please refer to Kumareswaran et al. 2012). Closed-loop models can be further distinguished as either fully closed-loop models and hybrid models.

In a fully closed-loop model, decisions for insulin dosing are exclusively based on parameters measured in the patient's body, e.g. blood glucose levels without knowledge of external information like food or exercise. In brief, changes in blood glucose levels would be measured and, based on these changes, the respective insulin dosage calculated and applied. Administration of insulin then affects the blood glucose level. Based on this feedback from the blood glucose level, the required level of insulin is again calculated and the next dose may be adjusted (figure 1). Such a system is reactive, meaning no anticipation and preemptive dosing based on experience or external information is possible. **A second major challenge for fully closed-loop models is to accommodate situations in which the blood glucose level changes rapidly, such as after meals or exercise. The human brain receives preliminary information even before spontaneous situations, for example thinking "I'm**

buying an ice-cream” before that meal is taken in, which is missing in a closed-loop artificial pancreas system. The conventional assumption is that the real pancreas works in a closed-loop system without this knowledge as well. However, recent work suggests that there’s a tighter link between brain and pancreatic function .

1. Hybrid models using both closed-loop control and external information, e.g. of meal size (figure 1), have therefore been proposed. These hybrid models reflect the proposed hypothalamic-pancreatic-relationship by including preliminary information about a future meal intake and might therefore be closer to the real pancreas. However, thinking “I’m buying an ice-cream” does not inform the pancreas about the exact composition of the meal; it may merely prepare the pancreas to respond to food intake. An artificial pancreas is different from its biological counterpart and its response time is mainly determined by the diffusion rate of injected insulin, therefore we propose investigating whether the hybrid model might, in future, be replaced by a fully closed-loop model to reduce the required frequency of user interventions. In a recent study, Weinzimer and coworkers compared fully closed-loop control with their hybrid or extended closed-loop control, in which an additional premeal bolus was used. The hybrid model reduced hyperglycemia after meals without inducing hypoglycemia , and the same result was seen in a similar model tested by another group in hospital settings . In both studies, patients were already in excellent glycemic control, and thus hypoglycemia was neither observed with the hybrid, nor the fully closed-loop control model . However, these hybrid models still need to be tested for control of potentially hypoglycemic conditions.

Despite the good performance of hybrid closed-loop model with external information, the overall aim is to develop a fully closed-loop model to obviate the scheduled lifestyle of diabetes patients in regard of meals and exercise. One way to achieve this is to develop an

algorithm that automatically detects meals by checking the blood glucose curve and either advises the patient or the automated insulin pump to apply an insulin bolus. A first attempt to develop an analytic model that, at least partially, anticipates fluctuations in glycemia, was performed by development of algorithms for meal detection or meal size estimation . When compared *in silico* to closed-loop control without information on meals, the meal size estimation algorithm improved the time spent in normoglycaemic range and even reduced the HbA_{1c} from 7.15 % (treatment without algorithm) to 6.43% (treatment with algorithm) in adolescents and from 6.69 % to 6.23 % in adults . However, long term studies should be performed in various settings to detect whether the addition of complexity to the glucose control system is worth the effort. Furthermore, meal detection is only one of many challenges for closed-loop systems of insulin delivery control.

Components of an artificial pancreas

Independent of the control model used, the components of an artificial pancreas system all bear their own challenges. These are either derived from technical aspects or based on the physiology of glucose regulation and have to be kept in mind during the development of control algorithms for insulin dose calculation.

Continuous glucose monitoring

Measuring the blood glucose level provides the minimal requirement for calculation of the insulin dose. Patients with type 1 diabetes are recommended to check their blood glucose levels at least three times a day . More frequent glucose measurements allow for the overall trend of the blood glucose level to be estimated, as opposed to isolated measurements with no

information on whether blood glucose levels are increasing or decreasing. As finger pricking can be painful, a less invasive method enabling the patient to perform more frequent measurements is highly desirable. More frequent glucose measurements allow for the overall trend of the blood glucose level to be estimated, as opposed to isolated measurements with no information on whether blood glucose levels are increasing or decreasing .

Among various CGM sensors (for review on CGM sensors please refer to Vaddiraju et al. 2012), subcutaneously implantable sensors are most comfortable for the patient. Due to an evaluation roadmap for CGM sensors presented by the Clinical and Laboratory Standards Institute , the most important challenges for CGM devices to date relate to the accuracy of the measurement, as well as to the real-time assessment.

The accuracy of subcutaneous glucose measurement devices has been subject of a long-lasting debate. In such devices, the sensor detects glucose in the interstitial fluid during its diffusion between the capillary and the target cell . Under steady-state conditions, interstitial glucose levels have been shown to be similar, but not precisely equal to, venous blood glucose levels in healthy individuals or animals . Rapid changes in blood glucose concentrations have been reported to affect the accuracy of the interstitial glucose sensing, namely causing the sensor to report glucose levels below their actual values . Implanted intravenous glucose sensors, which would provide similar comfort, as well as faster and more accurate blood glucose measurements, are currently under investigation . However, to date subcutaneous measurement is still preferred due to lower risk of thrombosis and intravascular infection.

Real-time assessment of subcutaneous glucose measurement devices describes the lag between measurement of the glucose level by the sensor and the time at which the blood level insulin, delivered in response, reaches its maximum. A large delay reduces the ability of the system to respond to glucose levels in real-time and therefore its flexibility. Figure 2

summarizes the sequence of events, from a change in blood glucose in the body to the maximum effect of the insulin administered in response . First, changes in blood glucose levels are mirrored by the interstitial blood glucose levels after a 5 – 10 min delay . Measurement of those interstitial glucose levels is commonly performed either by electrochemical sensors or microdialysis techniques , which both take another 3 – 12 min . Next, the digital filtering of the glucose measurement can take another 1 – 2 minutes, and is required to compensate for background noise. At this point, the algorithm will take some time to calculate the correct dose, but it is expected to be comparatively rapid. Finally, after insulin application, there is a delay before insulin becomes fully active in the blood. The latter depends on the type of insulin analog used, the total insulin dose and the individual pharmacokinetic parameters of the patient. Taken together, this can introduce a time delay between changes in blood glucose and insulin effect of up to 1 hour, which any dosage calculation will have to accommodate.

Insulin administration

Since the first studies on insulin treatment of patients with diabetes , which have led to the Nobel prize for Frederick Banting and Charles Best, insulin delivery devices have seen a number of development phases that have improved their performance and ease of use. For example, the time delay between application of insulin and the maximum plasma insulin concentration mentioned above have already been shortened by the availability of rapid-acting insulins. Furthermore, other administration routes also bear the potential to decrease the time until the maximum effect of insulin occurs.

Intravenous insulin application, which would enable the fastest insulin effect, exhibits certain limitations due to catheter complications . Intraperitoneal insulin is difficult to administer ,

and the availability of intraperitoneal insulin devices is currently limited. Inhalation might provide a novel application route for insulin, and suitable medical devices have been approved by the Food and Drug Administration in 2008 . But the bioavailability of inhaled insulin is less than in a subcutaneous application and is extremely variable in smokers or patients with a cold . Pharmaceuticals for oral delivery of insulin are currently under development but are far from routine clinical use.

Wide availability and ease of management are the major advantages of subcutaneous insulin administration . This currently makes the subcutaneous application route the most appropriate for routine injections and thus also for an artificial pancreas. When CGM is also performed subcutaneously, the system is commonly referred to as subcutaneous-subcutaneous (sc-sc) systems.

Existing algorithms

The development of algorithms for closed-loop calculation of insulin dosage is intensively investigated. The major candidates for such algorithms proposed in recent years use model predictive control (MPC) or the proportional integrative derivate (PID) methods . Current MPC systems require a model (typically a dynamical systems model) that can predict future glucose levels given known values for current glucose, insulin delivery and food intake. Such control then calculates the appropriate insulin infusion rate by minimizing the difference between the model-predicted glucose concentration and the target glucose level over a prediction time-window The duration of this time-window is chosen as the time in which the bulk of the effect is seen from the insulin or insulin analogue used. In the MPC literature, there is no reported way to balance the rapidity of a return to normoglycaemia against the amount of insulin delivered. PID systems consist of three components: the proportional (in

case of diabetes the difference between the actual glucose level and the desired glucose level – the error), the integral (accumulation of past errors over time) and the derivative (the rate of change of these errors) . In short, the PID algorithm estimates the required control (in case of diabetes the required delivery of insulin) based on a weighted sum of PID terms, in order to minimize these errors and so bring the system to the desired glucose level . Thus, the PID algorithm is rather reactive and lacks theoretical validation, whereas MPC algorithms represent a more proactive approach but require a good model of the dynamics.

To date, there is only one closed-loop algorithm commercially available, i.e. the B.Braun Space GlucoseControl (B. Braun, Melsungen, Germany). This algorithm is based on model predictive control and is provided for use in insulin treatment of critically ill patients in intensive care units (ICU). Such patients often develop peripheral insulin resistance and relative insulin deficiency with resulting hyperglycemia . Among others, this endocrine paradigm leads to increased gluconeogenesis from the body's stores and reduced glucose uptake and utilization. The former notion that resulting hyperglycemia would redistribute glucose towards organs that rely on glucose as fuel and, consecutively, improve the chance to survive was disapproved by evidence . Indeed, dysregulations of glycemia are associated with a negative outcome. However, bringing glucose back to normoglycemia in the critically ill by insulin infusion has shown to be a double-edged sword, since hypoglycemia and fluctuations of glycemia offset beneficial effects of glucose control when the target range is set too low . Therefore, critical care societies recommend controlling glycemia below 145 or 180 mg/dl, respectively . Glycemic control by the one commercial algorithm initially required an hourly measurement of glucose , but recent adaptation of the algorithm has derestricted this constraint . This algorithm is currently only used for critically ill patients and is applicable in the special setting of intensive care only. Though integrated closed-loop glucose control is very promising in clinical settings , algorithms for home patients might raise the concern of

missing control by human sense. This can be overcome by including an optional user-check of the insulin dosage values calculated by the algorithm, for example by providing the algorithm as a smartphone application as recently reported by Cobelli and coworkers . Despite the potential inclusion of a user-check, further improvements of the flexibility of an insulin calculating algorithm for independent outpatient use are highly desirable.

Possibilities to improve accuracy and flexibility of algorithms for artificial pancreas

In order to develop an algorithm appropriate for an outpatient's artificial pancreas, a number of issues still need to be addressed including the response to meals, exercise, stress and sleep. To adequately deal with these situations, the algorithm either has to be explicitly developed to identify and respond to each situation separately, or its overall flexibility will have to improve, i.e. by more closely mimicking the physiological function of a working human pancreas. The former approach is not promising, and so in the following section we will discuss the potential for improving the flexibility of such an algorithm.

Use of additional diagnostic indicators

Glucose homeostasis is maintained by various parameters including glucagon, epinephrine, insulin and others . Some of these parameters may be useful indicators of food intake or stress levels for closed-loop control systems, obviating the need for external information. Incretins, for example, could be a useful indicator of food intake. Upon ingestion of a meal, incretins

such as the glucagon-like peptide 1 (GLP-1) and gastric inhibitory peptide (GIP) promote the first phase secretion of insulin in proportion of the glucose content of the meal . Measurement of active GLP-1 after meals could therefore indicate that glucose levels will soon be elevated and insulin dosing could be administered earlier, in anticipation of this. Unfortunately, GLP-1 is degraded by the dipeptidyl peptidase IV after approximately 2 min , making it difficult to detect. Present assays available for GLP-1 detection would have to be refined before routine measurement of GLP-1 could be used in this way.

After measuring a parameter like glucose or in the future GLP-1, the relationship between the parameter and the appropriate insulin dose response will have to be found. The first, original “minimal model” of the insulin-glucose relationship was based on mathematical models of glucose and insulin kinetics only , subsuming the regulatory roles of many organ systems in these two compartments. Instead, current research uses models of physiologically based pharmacokinetics-pharmacodynamics. In these models, each organ system is treated as a separate compartment . For example, the mixed meal model of Roy and Parker displays the absorption of the major compartments of a meal from the gut . These models are currently used for simulation of diabetes and evaluating algorithms *in silico*. However, they can also be used to predict the pharmacodynamic response to hypothetical dosing strategies, and may, therefore, be incorporated into a future algorithm.

Use of alternative or additional drugs regulating glucose homeostasis

Mimicking the physiological insulin pharmacokinetics in diabetes treatment is very demanding, especially for a dosage calculating algorithm. In a healthy pancreatic beta-cell, an increased blood glucose level and therefore the stimulation of the beta-cell by glucose induces a biphasic insulin release . Due to the increment in plasma glucose levels, a rapid peak of

insulin secretion is followed by a slowly increasing second phase of insulin secretion . The peak phase of insulin secretion is due to pre-formed insulin stored in mature vesicles and is thought to suppress the hepatic glucose output . The second phase requires new synthesis of proteins and increases slowly until the cell is adapted or the glucose stimulation ends. This biphasic profile of insulin secretion presents the researchers with the need of implicating the pharmacokinetics of different analogs of insulin into the calculations of the control algorithm. The time delay between application of insulin and the maximum plasma insulin concentration mentioned above have already been shortened by the availability of rapid-acting insulins (for review on insulin analogs see . However, the fact that minimizing this time delay would markedly help to ensure normoglycaemia in closed-loop models of artificial pancreas has raised the request for ultra-rapid-acting insulins. To achieve such a biphasic profile of insulin secretion, computational scientists need to include the pharmacokinetics of different analogs of insulin into the calculations of the control algorithm.

Another key player in glucose homeostasis is glucagon. Its role has been included in recent dynamic models and inclusion of such models in the artificial pancreas would more closely mimic the physiological pancreas reactions and thereby increase the flexibility and accuracy of the system. In a healthy pancreas, glucagon counters the effects of insulin, thus leading to elevation of blood glucose levels. Recently developed artificial pancreas systems apply pumps capable of both insulin and glucagon application . Two independent research groups reported glucagon treatment to prevent and to reverse hypoglycemia in bihormonal closed-loop systems . Minimizing glucagon dosage in two studies avoided side effects like nausea or gastrointestinal discomfort , though long term studies remain to be conducted. Thus, glucagon treatment represents a promising option and should be considered in future development of control algorithms for artificial pancreas.

In addition, recent research has focused on the usability of new drugs regulating the glucose homeostasis. Amylin, for example, is a peptide hormone co-secreted with insulin by the pancreatic beta cells which has similar functions to insulin . Application of amylin was successfully tested in clinical trials and even in closed-loop systems of combined insulin and amylin delivery . In addition, the peptide GLP-1 could not only be measured for prediction of the correct insulin dose but may also be administered in addition to insulin . However, exenatide, the first FDA approved GLP-1 agonist, is still assigned with a safety alert and could therefore not yet be recommended for routine use.

All these modifications mentioned might increase the flexibility of the artificial pancreas systems, but may also increase the complexity of the algorithms used. This issue might be addressed by using learning algorithms, which allow for dosing control to be flexibly optimized with respect to the biological system.

Use of adaptive algorithms

The complexity of insulin delivery and the demanding goal of maintaining normoglycemia necessitate a complex, adaptive and flexible algorithm, which may be achieved with the use of machine learning techniques as shown in some approaches . Appropriate machine learning algorithms are able to analyze training data, recognize complex patterns and on the basis of such patterns apply the knowledge to other data to predict their behavior . The principle of a learning algorithm is depicted in figure 3.

There are three general differences between traditional PID and MPC algorithms used for diabetes so far and machine learning approaches . First, machine learning is based on recognition of patterns instead of implication of defined hypotheses. It improves the accuracy

of the system, because it includes initially unidentified variables that might be overlooked by the traditional hypothesis based systems . A second advantage of machine learning approaches is that they consider interactions between variables instead of minimizing or ignoring them as traditional models do . This might result in more complex models, but the challenges resulting from glucose homeostasis mentioned above justify their use for diabetes treatment. Third, machine learning approaches imply the risk of developing a model which can perform perfectly on the training data without generalizing well to unseen data (called *over-fitting*) . It is therefore important to correct for *over-fitting* by using cross-validation and regularization techniques .

The first attempt towards including machine learning algorithms into diabetes care was the use of supervised learning with artificial neural network (ANN) classification for diabetes treatment . ANNs algorithms infer a function minimizing the error between calculated parameters and desired parameters with the help of supervised / labeled training data. Of note, in supervised learning, the labeling of data needs to be performed by an expert, which is time consuming and is prone to human error. In terms of blood glucose level prediction and insulin regimen recommendations, ANNs work well in short-term predictions even in closed-loop systems . However, they have not yet been tested for long-term predictions of the blood glucose level. Supervised learning systems, such as that used by Robertson and colleagues, need good training data which include the desired response . These data can be expensive and/or time-consuming to collect, they assume that good responses are known, and in general only lead to reliable predictions for situations similar to those in the training data . Furthermore, glucose control in diabetes patients is an on-going task requiring regular control responses. The static input-output nature of supervised learning ignores this and therefore errors might propagate.

Reinforcement learning algorithms – learning to maintain normoglycemia?

Another potential approach to shape diabetes treatment is the use of reinforcement learning (RL) algorithms. RL is a branch of machine learning, concerned with how an agent chooses actions to control a system. It is suited to problems including sequences of decisions along a time-line. Additionally, it can be used when decisions depend on the observed state, where effects may be remote in time from actions that induce them, and where there is some notion of preferred state(s) for the system. This is true for the artificial pancreas system, as there is a need to continuously observe the patients glucose level and determine the ideal time and amount for insulin delivery. Moreover, RL can be performed directly on real data, or it can interact with a dynamical system represented by a mathematical model and in general it makes only very modest assumptions about this system . A broad classification of the types of control algorithms in terms of performance (glucose control, amount of delivered insulin, reaction time of the system) or personalization is shown in figure 4.

The principle of RL is based on the interaction between a decision-making and self-learning agent and its environment. At each time point, the *agent* chooses an *action* to modify the environment. The environment changes its *state* and sends this information and a *numerical reward* according to the previous action back to the agent. Mapping of a particular state to a certain action is called policy of the algorithm and defines the behavior of the agent at each time step . The goal of RL is to learn an optimal policy and thus maximize the amount of reward it receives over time . To achieve that goal, the agent should not only choose the action which brings the most reward in one run (exploitation), it should also consider other possibilities to increase the overall reward (exploration). A balancing between exploitation and exploration is needed to generalize from experience. The agent need to *explore* unusual

states in the system by occasionally choosing unpromising actions during the learning procedure, in order to choose good actions for even these rare states during normal operation. The procedure of a RL algorithm for diabetes is depicted in figure 5.

In comparison to other traditional control strategies, RL does not require a detailed description of the environment in terms of a well-represented model or labeled training data as in supervised learning strategies. After a learning procedure, the agent develops a policy and thus a control strategy from experience to predict certain situations and rewards without a necessary mathematical specification of the environment. Another advantage of RL algorithms is that they are uniquely suited to systems with inherent time delays as these are present due to the subcutaneous glucose measurements and insulin injections. RL can also be used with large or even infinite state sets, which makes that approach useful for the different glycemic levels that occur during continuous glucose measuring.

Two potential criticisms of conventional RL methods are relevant to the case of insulin delivery control: The learned control is *black-box*, meaning it cannot be readily reused or generalized from, and they are not very efficient in terms of data. The efficacy issue arises, because conventional RL algorithms do not build explicit models of the environment, and are therefore sometimes referred to as model-free RL. To address this, we recommend the use of one or both of the following techniques: model-based or data-efficient RL. Model-based RL builds a dynamical model of the control problem through experience and uses this model to train an on-board model-free RL algorithm, e.g. Dyna-Q . Here, **we define a model-based reinforcement learning algorithm as one that maintains a system model, which is updated *on-line* – e.g. from real data as it is observed, and which optimizes a reinforcement learner using this model. This model can be either entirely constructed from empirical data, or can use prior knowledge**

to constrain the family of models considered. For our insulin delivery system, a model-based reinforcement learner could therefore define *a priori* a dynamical system structure where blood-glucose depends on insulin and beta-cells in a specified way, but use real data to update the parameters of that model. Data-efficient RL algorithms focus on making the most efficient use of experience gained so far, e.g. fitted-Q . The latter approach has been proposed for use in clinical domains , and recent work, which combines this with a model-based RL approach, has shown remarkable data-efficiency in robotics tasks

To date, RL algorithms have been proposed for the treatment of epilepsy , renal anaemia or the control of anesthesia . In the case of renal anaemia, the RL algorithm was informed by an MPC, showing that these two approaches do not necessarily exclude each other but can be used in parallel. When used in closed-loop control of anesthesia *in silico*, a RL algorithm outperforms a PID control algorithm by less overshoot of the depth of hypnosis and faster achievement of steady state . Thus, use of the RL algorithm in this closed-loop control setting resulted in tighter control, a principle that could be administrable for patients with diabetes. An initial study on using a RL algorithm to control an artificial pancreas, reported good performance in controlling hyperglycemia *in silico* . In this study, the *state* was defined as different glycemetic ranges, *action* was defined as insulin infusion and *reward* was set equal to the difference of the glucose concentration from its target value . *In silico* application of the algorithm led to correction of hyperglycemia to normoglycemia. **However, given the criticism on RL mentioned above it is important to point out that this research used model-free RL in an off-line manner on a fixed model with fixed parameters, which is distinct from a model-based reinforcement learner that updates its model *on-line*. There is no description of how to verify the accuracy of the model or how to adapt the controller to individual patients, although the authors do acknowledge that these are research issues.** Furthermore, the tests included *in*

in silico patients only and no *in vivo* studies were performed. The same is true for more recent work using an actor-critic reinforcement learning algorithm, which shows promising results in adults and children *in silico*, which still has to be verified in a clinical trial. Moreover, the authors do not discuss how the early exploratory phase of the algorithm can be safely achieved *in vivo*. Thus, only a first proof of principle for the usability of RL algorithms in diabetes has been successfully performed. It would be worth comparing such RL algorithms to other algorithms in a larger setting in future studies and to further exploit their full potential in terms of flexible reactions on changes in the blood glucose levels of diabetes patients.

Expert commentary

Algorithms for closed-loop models of insulin treatment have to deal with demanding challenges due to the complex physiology of glucose homeostasis as well as technical limitations of the components of an artificial pancreas. The flexible reactivity especially required for outpatients suggests the use of data-driven machine learning algorithms. Among those, RL algorithms exhibit a great potential to deal with the time delay produced by the CGM system. In view of the available evidence, it can be summarized that RL algorithms provide a very promising approach for flexibly and independently maintaining normoglycemia in artificial pancreas systems. **To date, the vast majority of papers reporting the development of algorithms demonstrate control in limited scenarios *in silico*, e.g. an insulin spike after a single meal. We assert that stochastic models are essential to assess the reliability and stability of an algorithm for periods containing multiple meal events, whereas future *in vivo* studies of closed-loop algorithms are required to reliably assess performance and personalization.**

Five-year view

In the future, minimization of the time delay between changes in the glucose level and the full effect of insulin as well as maximization of the accuracy of the subcutaneous glucose measuring devices will be the subject of studies on the components of artificial pancreas. Ultra-rapid acting insulins are under development as well as substances like GLP-1 or amylin, increasing the flexibility of glucose regulation. However, the more other substances for regulation of blood glucose come up, the more individual and flexible becomes glucose control - and the more complex. This increases the need for smart, personalized algorithms calculating insulin delivery in the future and might amongst others be achieved by the use of reinforcement learning algorithms.

Key issues

- Maintaining normoglycemia is crucial in patients with diabetes mellitus or severe illness and is usually achieved by administration of insulin.
- An artificial pancreas system for closed-loop insulin delivery consists of a continuous glucose monitoring device, an algorithm calculating the correct amount of insulin and a pump delivering insulin.
- Current challenges for calculation of the correct dose include technical issues most notably with regard to the time delay between changes in the glucose level and the maximum effect of insulin.
- Upcoming substances for glucose regulation like glucagon or amylin increase both the flexibility and the complexity of the system.

- The individualized treatment regimes and the complex glucose regulating parameters elevate the need for smart, flexible algorithms calculating the insulin dose.
- Algorithms used in the past were initially based on Model Predictive Control or Proportional Integral Derivative Control.
- Machine learning algorithms and especially reinforcement learning algorithms provide the advantages to learn the individual glucose pattern of a diabetic patient in spite of a time delay and to handle complex and external information to provide adaptive drug delivery after a learning procedure.
- For machine learning approaches, care must be taken to acquire appropriate data for the learning phase whereby the data should be representative, sufficient and optimally noise-reduced.
- To maximise the effectiveness of data driven approaches, cross-validation and regularization techniques should be used and an extensive testing phase has to be performed.

Acknowledgements/Author contributions

M.B. wrote the manuscript. L.D., K.R., A.T., B.E., M.W. and A.F. reviewed and edited the manuscript. Martin Westphal, Arn Tellmann, Katrin Reichel and Melanie Bothe are employees of Fresenius Kabi Deutschland GmbH. Fresenius Kabi Deutschland GmbH distributes pumps for insulin delivery. All authors declare having no relationship with other companies that make products relevant to the paper and have no conflict of interest with the present work.

Figures

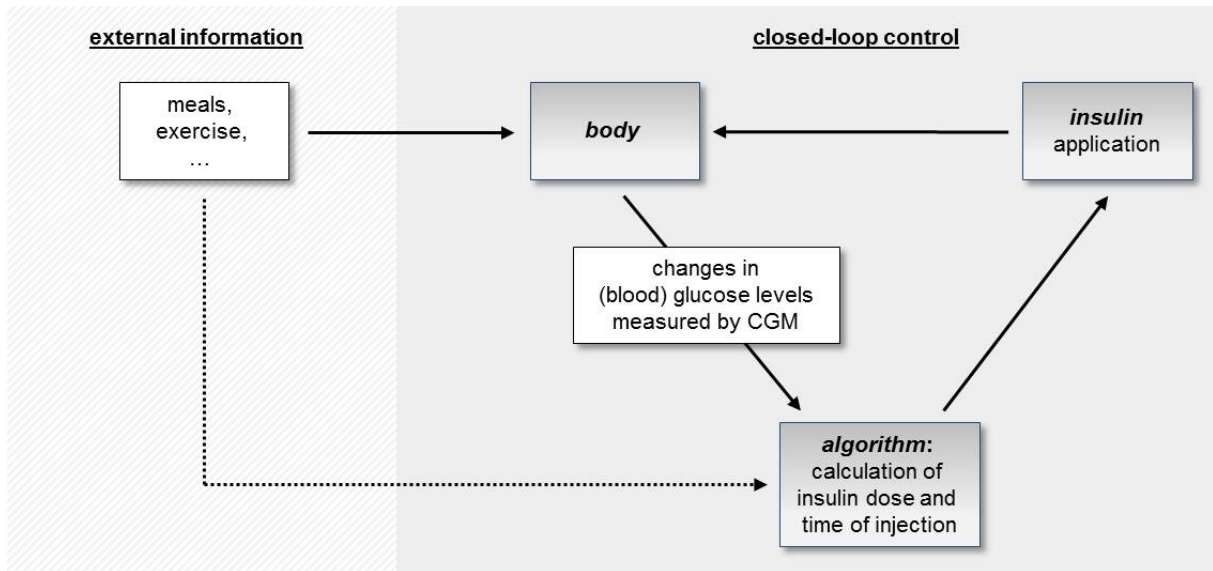


Figure 1: Closed-loop control model for insulin delivery with (hybrid model, dotted line) and without external information. In the latter case, the algorithm reacts directly on the changes in glucose levels evoked by meals or exercise without getting external information.

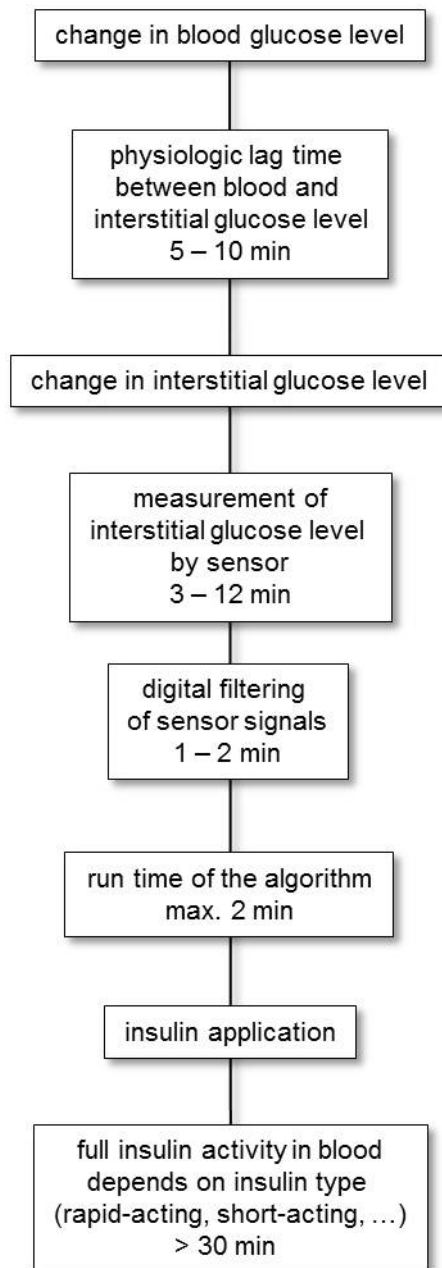


Figure 2: Components of the time delay between blood glucose level and maximum blood insulin level leading to a lag time of approximately 1h.

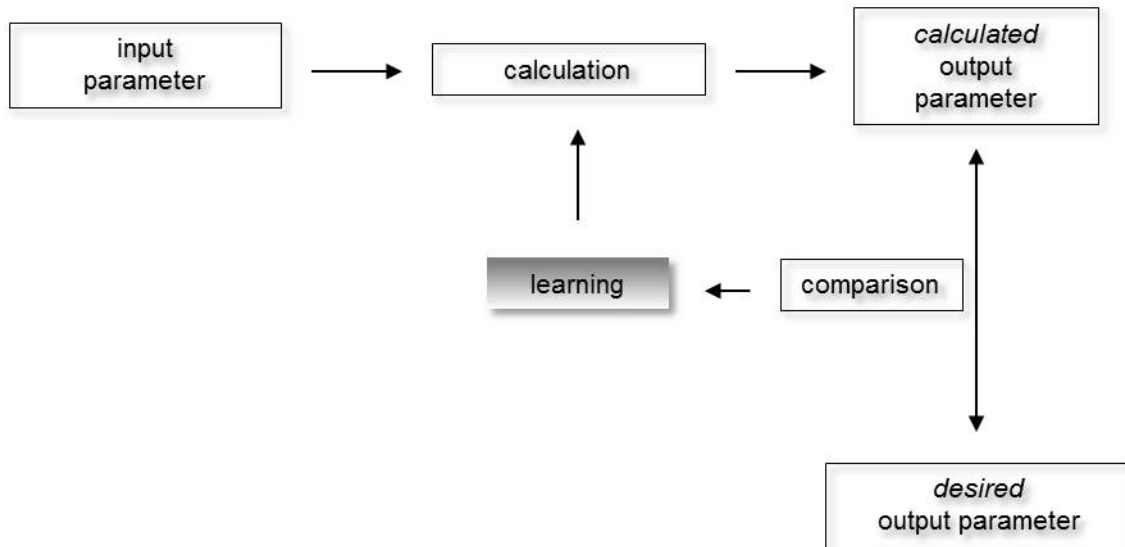


Figure 3: Principle of a learning algorithm. Comparison of the calculated output parameter with the desired output parameter leads to learning of the algorithm.

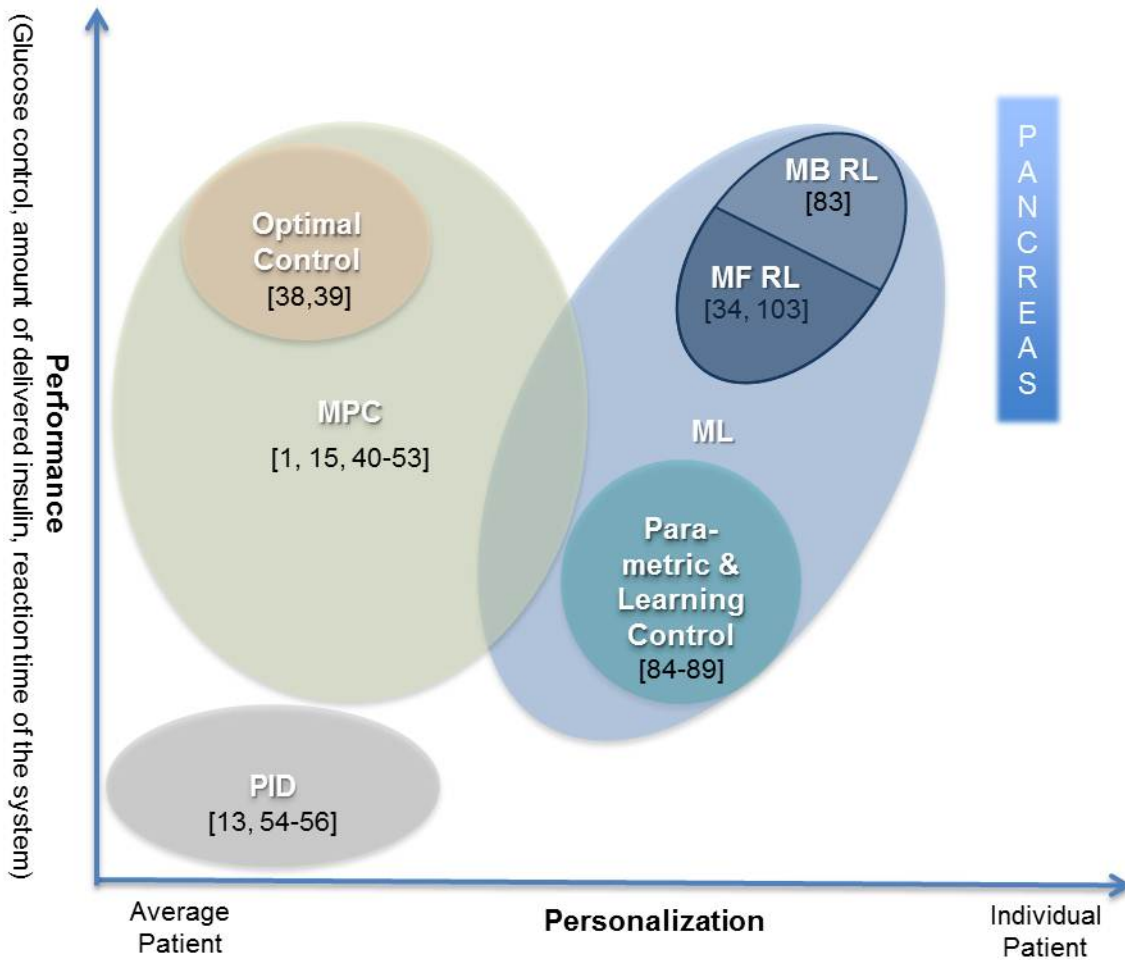


Figure 4: Schematic overview of commonly used types of algorithms for glucose control as well as machine learning and reinforcement learning algorithms. The references are to be seen as examples without valuation and the list is not intended to be exhaustive. MB RL = Model-Based Reinforcement Learning; MF RL = Model-Free Reinforcement Learning; ML = Machine Learning; MPC = Model Predictive Control; PID = Proportional Integral Derivative Control; RL = Reinforcement Learning;

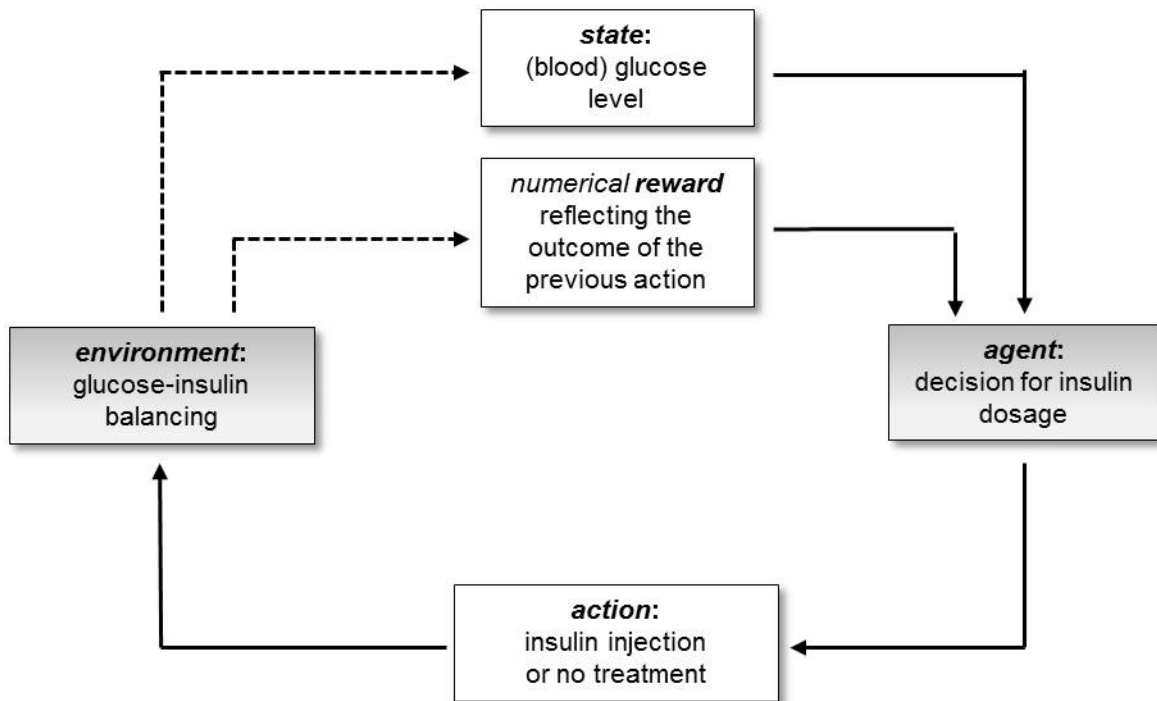


Figure 5: Reinforcement learning algorithms for diabetes. Changes in the state lead to an action of the agent, which changes the environment. The agent receives a numerical reward from the environment, which together with the next status will influence the next action.

References

1. Pachler C, Plank J, Weinhandl H *et al.* Tight glycaemic control by an automated algorithm with time-variant sampling in medical ICU patients. *Intensive care medicine*, 34(7), 1224-1230 (2008).
2. Elleri D, Allen JM, Biagioni M *et al.* Evaluation of a portable ambulatory prototype for automated overnight closed-loop insulin delivery in young people with type 1 diabetes. *Pediatric diabetes*, 13(6), 449-453 (2012).
3. Atlas E, Nimri R, Miller S, Grunberg EA, Phillip M. MD-logic artificial pancreas system: a pilot study in adults with type 1 diabetes. *Diabetes care*, 33(5), 1072-1076 (2010).
4. American Diabetes Association. Standards of medical care in diabetes--2011. *Diabetes care*, 34 Suppl 1, S11-61 (2011).
5. Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *The New England journal of medicine*, 329(14), 977-986 (1993).
6. International Diabetes Federation. IDF Diabetes Atlas, 5th edn. . (Ed.^(Eds) (Brussels, Belgium, 2011)
7. Clemens AH, Chang PH, Myers RW. The development of Biostator, a Glucose Controlled Insulin Infusion System (GCIIS). *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme*, Suppl 7, 23-33 (1977).
8. Farmer TG, Jr., Edgar TF, Peppas NA. The future of open- and closed-loop insulin delivery systems. *The Journal of pharmacy and pharmacology*, 60(1), 1-13 (2008).
9. Kumareswaran K, Evans ML, Hovorka R. Closed-loop insulin delivery: towards improved diabetes care. *Discovery medicine*, 13(69), 159-170 (2012).
10. Thorens B. Brain glucose sensing and neural regulation of insulin and glucagon secretion. *Diabetes, obesity & metabolism*, 13 Suppl 1, 82-88 (2011).

11. Thorens B. Sensing of glucose in the brain. *Handbook of experimental pharmacology*, (209), 277-294 (2012).
12. Weinzimer SA, Steil GM, Swan KL, Dziura J, Kurtz N, Tamborlane WV. Fully automated closed-loop insulin delivery versus semiautomated hybrid control in pediatric patients with type 1 diabetes using an artificial pancreas. *Diabetes care*, 31(5), 934-939 (2008).
13. Renard E, Place J, Cantwell M, Chevassus H, Palerm CC. Closed-loop insulin delivery using a subcutaneous glucose sensor and intraperitoneal insulin delivery: feasibility study testing a new model for the artificial pancreas. *Diabetes care*, 33(1), 121-127 (2010).
14. Dassau E, Bequette BW, Buckingham BA, Doyle FJ, 3rd. Detection of a meal using continuous glucose monitoring: implications for an artificial beta-cell. *Diabetes care*, 31(2), 295-300 (2008).
15. Lee H, Buckingham BA, Wilson DM, Bequette BW. A closed-loop artificial pancreas using model predictive control and a sliding meal size estimator. *Journal of diabetes science and technology*, 3(5), 1082-1090 (2009).
16. Vaddiraju S, Burgess DJ, Tomazos I, Jain FC, Papadimitrakopoulos F. Technologies for continuous glucose monitoring: current problems and future promises. *Journal of diabetes science and technology*, 4(6), 1540-1562 (2010).
17. Clinical and Laboratory Standards Institute. Performance Metrics for Continuous Intersitital Glucose Monitoring; Approved Guideline. (Ed.^(Eds) (2009)
18. Cengiz E, Tamborlane WV. A tale of two compartments: interstitial versus blood glucose monitoring. *Diabetes technology & therapeutics*, 11 Suppl 1, S11-16 (2009).
19. Aussedat B, Dupire-Angel M, Gifford R, Klein JC, Wilson GS, Reach G. Interstitial glucose concentration and glycemia: implications for continuous subcutaneous glucose monitoring. *American journal of physiology. Endocrinology and metabolism*, 278(4), E716-728 (2000).

20. Lonroth P, Jansson PA, Smith U. A microdialysis method allowing characterization of intercellular water space in humans. *The American journal of physiology*, 253(2 Pt 1), E228-231 (1987).
21. Rebrin K, Steil GM, van Antwerp WP, Mastrototaro JJ. Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring. *The American journal of physiology*, 277(3 Pt 1), E561-571 (1999).
22. Baek YH, Jin HY, Lee KA *et al.* The Correlation and Accuracy of Glucose Levels between Interstitial Fluid and Venous Plasma by Continuous Glucose Monitoring System. *Korean diabetes journal*, 34(6), 350-358 (2010).
23. Kulcu E, Tamada JA, Reach G, Potts RO, Lesho MJ. Physiological differences between interstitial glucose and blood glucose measured in human subjects. *Diabetes care*, 26(8), 2405-2409 (2003).
24. Beier B, Musick K, Matsumoto A, Panitch A, Nauman E, Irazoqui P. Toward a continuous intravascular glucose monitoring system. *Sensors*, 11(1), 409-424 (2011).
25. Skjaervold NK, Solligard E, Hjelme DR, Aadahl P. Continuous measurement of blood glucose: validation of a new intravascular sensor. *Anesthesiology*, 114(1), 120-125 (2011).
26. Aye T, Block J, Buckingham B. Toward closing the loop: an update on insulin pumps and continuous glucose monitoring systems. *Endocrinology and metabolism clinics of North America*, 39(3), 609-624 (2010).
27. Keenan DB, Mastrototaro JJ, Voskanyan G, Steil GM. Delays in minimally invasive continuous glucose monitoring devices: a review of current technology. *Journal of diabetes science and technology*, 3(5), 1207-1214 (2009).
28. Rebrin K, Sheppard NF, Jr., Steil GM. Use of subcutaneous interstitial fluid glucose to estimate blood glucose: revisiting delay and sensor offset. *Journal of diabetes science and technology*, 4(5), 1087-1098 (2010).
29. Mastrototaro J. The MiniMed Continuous Glucose Monitoring System (CGMS). *Journal of pediatric endocrinology & metabolism : JPEM*, 12 Suppl 3, 751-758 (1999).

30. Glucose Monitoring Study Group. Continuous glucose monitoring by means of the microdialysis technique: underlying fundamental aspects. *Diabetes technology & therapeutics*, 5(4), 545-561 (2003).
31. Banting FG, Campbell WR, Fletcher AA. Further Clinical Experience with Insulin (Pancreatic Extracts) in the Treatment of Diabetes Mellitus. *British medical journal*, 1(3236), 8-12 (1923).
32. Gilchrist JA, Best CH, Banting FG. Observations with Insulin on Department of Soldiers' Civil Re-Establishment Diabetics. *Canadian Medical Association journal*, 13(8), 565-572 (1923).
33. Renard E. Insulin delivery route for the artificial pancreas: subcutaneous, intraperitoneal, or intravenous? Pros and cons. *Journal of diabetes science and technology*, 2(4), 735-738 (2008).
34. Yasini S, Naghibi-Sistani MB, Karimpour A. Agent-based Simulation for Blood Glucose Control in Diabetic Patients. *World Academy of Science, Engineering and Technology*, 57(33), 11 (2009).
35. McMahon GT, Arky RA. Inhaled insulin for diabetes mellitus. *The New England journal of medicine*, 356(5), 497-502 (2007).
36. Kapitza C, Zijlstra E, Heinemann L, Castelli MC, Riley G, Heise T. Oral insulin: a comparison with subcutaneous regular human insulin in patients with type 2 diabetes. *Diabetes care*, 33(6), 1288-1290 (2010).
37. Najafzadeh H, Kooshapur H, Kianidehkordi F. Evaluation of an oral insulin formulation in normal and diabetic rats. *Indian journal of pharmacology*, 44(1), 103-105 (2012).
38. Acikgoz US, Diwekar UM. Blood glucose regulation with stochastic optimal control for insulin-dependent diabetic patients. *Chemical Engineering Science*, 65(3), 1227–1236 (2010).
39. Chee F, Savkin AV, Fernando TL, Nahavandi S. Optimal H infinity insulin injection control for blood glucose regulation in diabetic patients. *IEEE transactions on bio-medical engineering*, 52(10), 1625-1631 (2005).

40. Bruttomesso D, Farret A, Costa S *et al.* Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: preliminary studies in Padova and Montpellier. *Journal of diabetes science and technology*, 3(5), 1014-1021 (2009).
41. Clarke WL, Anderson S, Breton M, Patek S, Kashmer L, Kovatchev B. Closed-loop artificial pancreas using subcutaneous glucose sensing and insulin delivery and a model predictive control algorithm: the Virginia experience. *Journal of diabetes science and technology*, 3(5), 1031-1038 (2009).
42. Elleri D, Allen JM, Nodale M *et al.* Automated overnight closed-loop glucose control in young children with type 1 diabetes. *Diabetes technology & therapeutics*, 13(4), 419-424 (2011).
43. Ellingsen C, Dassau E, Zisser H *et al.* Safety constraints in an artificial pancreatic beta cell: an implementation of model predictive control with insulin on board. *Journal of diabetes science and technology*, 3(3), 536-544 (2009).
44. Gillis R, Palerm CC, Zisser H, Jovanovic L, Seborg DE, Doyle FJ. Glucose estimation and prediction through meal responses using ambulatory subject data for advisory mode model predictive control. *Journal of diabetes science and technology*, 1(6), 825-833 (2007).
45. Grosman B, Dassau E, Zisser HC, Jovanovic L, Doyle FJ, 3rd. Zone model predictive control: a strategy to minimize hyper- and hypoglycemic events. *Journal of diabetes science and technology*, 4(4), 961-975 (2010).
46. Hovorka R, Canonico V, Chassin LJ *et al.* Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiological measurement*, 25(4), 905-920 (2004).
47. Kan S, Onodera H, Furutani E *et al.* Novel control system for blood glucose using a model predictive method. *ASAIO journal*, 46(6), 657-662 (2000).

48. Magni L, Forgione M, Toffanin C *et al.* Run-to-run tuning of model predictive control for type 1 diabetes subjects: in silico trial. *Journal of diabetes science and technology*, 3(5), 1091-1098 (2009).
49. Murphy HR, Elleri D, Allen JM *et al.* Closed-loop insulin delivery during pregnancy complicated by type 1 diabetes. *Diabetes care*, 34(2), 406-411 (2011).
50. Schaller HC, Schaupp L, Bodenlenz M *et al.* On-line adaptive algorithm with glucose prediction capacity for subcutaneous closed loop control of glucose: evaluation under fasting conditions in patients with Type 1 diabetes. *Diabetic medicine : a journal of the British Diabetic Association*, 23(1), 90-93 (2006).
51. Schlotthauer G, Gamero LG, Torres ME, Nicolini GA. Modeling, identification and nonlinear model predictive control of type I diabetic patient. *Medical engineering & physics*, 28(3), 240-250 (2006).
52. Wang Y, Dassau E, Doyle FJ, 3rd. Closed-loop control of artificial pancreatic Beta-cell in type 1 diabetes mellitus using model predictive iterative learning control. *IEEE transactions on bio-medical engineering*, 57(2), 211-219 (2010).
53. Wilinska ME, Budiman ES, Taub MB *et al.* Overnight closed-loop insulin delivery with model predictive control: assessment of hypoglycemia and hyperglycemia risk using simulation studies. *Journal of diabetes science and technology*, 3(5), 1109-1120 (2009).
54. Panteleon AE, Loutseiko M, Steil GM, Rebrin K. Evaluation of the effect of gain on the meal response of an automated closed-loop insulin delivery system. *Diabetes*, 55(7), 1995-2000 (2006).
55. Steil GM, Palerm CC, Kurtz N *et al.* The effect of insulin feedback on closed loop glucose control. *The Journal of clinical endocrinology and metabolism*, 96(5), 1402-1408 (2011).
56. Wintergerst KA, Deiss D, Buckingham B *et al.* Glucose control in pediatric intensive care unit patients using an insulin-glucose algorithm. *Diabetes technology & therapeutics*, 9(3), 211-222 (2007).

57. Cobelli C, Renard E, Kovatchev B. Artificial pancreas: past, present, future. *Diabetes*, 60(11), 2672-2682 (2011).
58. Hovorka R, Kumareswaran K, Harris J *et al.* Overnight closed loop insulin delivery (artificial pancreas) in adults with type 1 diabetes: crossover randomised controlled studies. *Bmj*, 342, d1855 (2011).
59. Elleri D, Dunger DB, Hovorka R. Closed-loop insulin delivery for treatment of type 1 diabetes. *BMC medicine*, 9, 120 (2011).
60. Hovorka R, Kremen J, Blaha J *et al.* Blood glucose control by a model predictive control algorithm with variable sampling rate versus a routine glucose management protocol in cardiac surgery patients: a randomized controlled trial. *The Journal of clinical endocrinology and metabolism*, 92(8), 2960-2964 (2007).
61. Egi M, Finfer S, Bellomo R. Glycemic control in the ICU. *Chest*, 140(1), 212-220 (2011).
62. Kavanagh BP. Glucose in the ICU - Evidence, Guidelines, and Outcomes. *The New England journal of medicine*, (2012).
63. Cordingley JJ, Vlasselaers D, Dormand NC *et al.* Intensive insulin therapy: enhanced Model Predictive Control algorithm versus standard care. *Intensive care medicine*, 35(1), 123-128 (2009).
64. Blaha J, Kopecky P, Matias M *et al.* Comparison of three protocols for tight glycemic control in cardiac surgery patients. *Diabetes care*, 32(5), 757-761 (2009).
65. Breton M, Farret A, Bruttomesso D *et al.* Fully Integrated Artificial Pancreas in Type 1 Diabetes: Modular Closed-Loop Glucose Control Maintains Near Normoglycemia. *Diabetes*, 61(9), 2230-2237 (2012).
66. Cobelli C, Renard E, Kovatchev BP *et al.* Pilot studies of wearable outpatient artificial pancreas in type 1 diabetes. *Diabetes care*, 35(9), e65-67 (2012).
67. Teixeira RE, Malin S. The next generation of artificial pancreas control algorithms. *Journal of diabetes science and technology*, 2(1), 105-112 (2008).

68. D'Alessio DA, Vahl TP. Glucagon-like peptide 1: evolution of an incretin into a treatment for diabetes. *American journal of physiology. Endocrinology and metabolism*, 286(6), E882-890 (2004).
69. Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *The Journal of clinical investigation*, 68(6), 1456-1467 (1981).
70. Roy A, Parker RS. Mixed meal modeling and disturbance rejection in type I diabetic patients. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 1, 323-326 (2006).
71. Steil GM, Panteleon AE, Rebrin K. Closed-loop insulin delivery-the path to physiological glucose control. *Advanced drug delivery reviews*, 56(2), 125-144 (2004).
72. Luzi L, DeFronzo RA. Effect of loss of first-phase insulin secretion on hepatic glucose production and tissue glucose disposal in humans. *The American journal of physiology*, 257(2 Pt 1), E241-246 (1989).
73. Hartman I. Insulin analogs: impact on treatment success, satisfaction, quality of life, and adherence. *Clinical medicine & research*, 6(2), 54-67 (2008).
74. Castle JR, Engle JM, El Youssef J *et al.* Novel use of glucagon in a closed-loop system for prevention of hypoglycemia in type 1 diabetes. *Diabetes care*, 33(6), 1282-1287 (2010).
75. El-Khatib FH, Russell SJ, Nathan DM, Sutherlin RG, Damiano ER. A bihormonal closed-loop artificial pancreas for type 1 diabetes. *Science translational medicine*, 2(27), 27ra27 (2010).
76. Adeghate E, Kalasz H. Amylin analogues in the treatment of diabetes mellitus: medicinal chemistry and structural basis of its function. *The open medicinal chemistry journal*, 5(Suppl 2), 78-81 (2011).

77. Schmitz O, Brock B, Rungby J. Amylin agonists: a novel approach in the treatment of diabetes. *Diabetes*, 53 Suppl 3, S233-238 (2004).
78. Heptulla RA, Rodriguez LM, Mason KJ, Haymond MW. Twenty-four-hour simultaneous subcutaneous Basal-bolus administration of insulin and amylin in adolescents with type 1 diabetes decreases postprandial hyperglycemia. *The Journal of clinical endocrinology and metabolism*, 94(5), 1608-1611 (2009).
79. Weinzimer SA, Sherr JL, Cengiz E *et al.* Effect of pramlintide on prandial glycemic excursions during closed-loop control in adolescents and young adults with type 1 diabetes. *Diabetes care*, 35(10), 1994-1999 (2012).
80. Macconell L, Brown C, Gurney K, Han J. Safety and tolerability of exenatide twice daily in patients with type 2 diabetes: integrated analysis of 5594 patients from 19 placebo-controlled and comparator-controlled clinical trials. *Diabetes, metabolic syndrome and obesity : targets and therapy*, 5, 29-41 (2012).
81. Nikfar S, Abdollahi M, Salari P. The efficacy and tolerability of exenatide in comparison to placebo; a systematic review and meta-analysis of randomized clinical trials. *Journal of pharmacy & pharmaceutical sciences : a publication of the Canadian Society for Pharmaceutical Sciences, Societe canadienne des sciences pharmaceutiques*, 15(1), 1-30 (2012).
82. U.S. Food and Drug Administration (FDA). Information for Healthcare Professionals: Reports of Altered Kidney Function in patients using Exenatide (Marketed as Byetta). (Ed.^(Eds) (2009)
83. Reichel K, Dickens L, Tellmann A, Bothe MK, Westphal M, Faisal AA. Machine Learning for Closed-loop Insulin Delivery: Model & in-vivo Studies. (Ed.^(Eds) (Conference: Bioengineering 12, 2012)
84. Dazzi D, Taddei F, Gavarini A, Uggeri E, Negro R, Pezzarossa A. The control of blood glucose in the critical diabetic patient: a neuro-fuzzy method. *Journal of diabetes and its complications*, 15(2), 80-87 (2001).

85. El-Jabali AK. Neural network modeling and control of type 1 diabetes mellitus. *Bioprocess and biosystems engineering*, 27(2), 75-79 (2005).
86. Gogou G, Maglaveras N, Ambrosiadou BV, Goulis D, Pappas C. A neural network approach in diabetes management by insulin administration. *Journal of medical systems*, 25(2), 119-131 (2001).
87. Mougiakakou SG, Nikita KS. A neural network approach for insulin regime and dose adjustment in type 1 diabetes. *Diabetes technology & therapeutics*, 2(3), 381-389 (2000).
88. Otto E, Semotok C, Andrysek J, Basir O. An intelligent diabetes software prototype: predicting blood glucose levels and recommending regimen changes. *Diabetes technology & therapeutics*, 2(4), 569-576 (2000).
89. Trajanoski Z, Regittnig W, Wach P. Simulation studies on neural predictive control of glucose using the subcutaneous route. *Computer methods and programs in biomedicine*, 56(2), 133-139 (1998).
90. Bishop CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, 2007).
91. Waljee AK, Higgins PD. Machine learning in medicine: a primer for physicians. *The American journal of gastroenterology*, 105(6), 1224-1226 (2010).
92. G. R, D. LE, W. S, D. H. Blood Glucose Prediction Using Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study. *Journal of Electrical and Computer Engineering*, Volume 2011 (2011) (2011).
93. Zarkogianni K, Vazeou A, Mougiakakou SG, Prountzou A, Nikita KS. An insulin infusion advisory system based on autotuning nonlinear model-predictive control. *IEEE transactions on bio-medical engineering*, 58(9), 2467-2477 (2011).
94. Sutton R.S. BAG. Reinforcement Learning: An Introduction. *MIT Press*, (1998).
95. Pineau J, Guez A, Vincent R, Panuccio G, Avoli M. Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach. *International journal of neural systems*, 19(4), 227-240 (2009).

96. Moore BL, Doufas AG, Pyeatt LD. Reinforcement learning: a novel method for optimal control of propofol-induced hypnosis. *Anesthesia and analgesia*, 112(2), 360-367 (2011).
97. Sutton RS. *Integrated architectures for learning, planning, and reacting based on approximating dynamic programming* (1990).
98. Ernst D, Geurts P, Wehenkel L. Tree-Based Batch Mode Reinforcement Learning. *J. Mach. Learn. Res.*, 6, 503--556 (2005).
99. Ernst D, Stan G-B, Goncalves J, Wehenkel L. *Clinical data based optimal STI strategies for HIV; a reinforcement learning approach* (2006).
100. Deisenroth MP, Rasmussen CE. *PILCO: A Model-Based and Data-Efficient Approach to Policy Search* (2011).
101. Gaweda AE, Muezzinoglu MK, Jacobs AA, Aronoff GR, Brier ME. Model predictive control with reinforcement learning for drug delivery in renal anemia management. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 1, 5177-5180 (2006).
102. Moore BL, Quasny TM, Doufas AG. Reinforcement learning versus proportional-integral-derivative control of hypnosis in a simulated intraoperative patient. *Anesthesia and analgesia*, 112(2), 350-359 (2011).
103. Daskalaki E, Diem P, Mougiakakou SG. An Actor-Critic based controller for glucose regulation in type 1 diabetes. *Computer methods and programs in biomedicine*, (2012).