Online Adaptive Personalization of Supervised Learning Models in Mobile Health Treatment Systems

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Jonathan L Hughes

May 2020

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: _____

This Dissertation has been read and approved by the examining committee:

Advisor: Marc Breton

Committee Member: Laura Barnes

Committee Member: Donald Brown

Committee Member: Daniel Cherñavvsky

Committee Member: Cody Fleming

Committee Member: Stephen Patek

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

May 2020

Abstract

There is often significant heterogeneity present in the context of systems engineering problems. This heterogeneity can limit the effectiveness of policies and models that are designed to operate at a coarse, population level when the actual point of intervention is at the level of the specific and varying subgroups or individuals constituting the population. Thus, methods of model personalization may be required to achieve desired outcomes. In this dissertation, we propose a means of rapid, online model personalization of decision rules based on statistical learning models, GMAdapt, which is informed by the context of decision support systems for the management of type 1 diabetes. To evaluate the effectiveness of this procedure, we performed experiments using both numerical simulations and retrospective data analysis based on real-world clinical trials conducted at the UVa Center for Diabetes Technology. In addition to the adaptation procedure itself, we present a simulation based methodology for deconfounding data to address the issue of intervention generated label noise. This method is evaluated *in silico* using the UVa/Padova type 1 diabetes simulator and compared against some alternative methodologies for creating end-to-end systems capable of adaptively learning personalized decision rules in spite of system generated interventions and resulting label noise.

Contents

Chapter 1: Introduction

1.1 Thesis Statement

The goal of this dissertation is to present a method for rapidly developing personalized models informed by the context of a decision support systems used in the management of type 1 diabetes. The presented research uses a systems engineering approach to create a methodology designed to overcome the constraints and confounding factors endemic to healthcare related systems, and applicable to other fields which face similar issues.

1.2 Overview

There have been many noteworthy mechanisms driving recent technological advances in the field of machine learning and data science: the exponential increase in cheaply available computing power, the availability of "Big" data sets, the rise of sophisticated mathematical modeling techniques, etc., which have enabled computer systems to accomplish complex tasks that were unmanageable even a decade ago. Despite these advances, the creation of an artificial general intelligence system, if possible at all, still lies in the future. Until such systems are created, it will be the role of engineers not only to produce better technology, but also to formulate the problems we encounter in a manner that allows for the best available technology to be effectively employed to solve them. It is not enough that a system works in principle, that the mathematics behind it are sound, or that it works within an environment that is so controlled that it fails to resemble the "real-world" field of application. The system must work within the context of the real-world uncertainty and the constraints that it will actually encounter, not only in deployment stage but even throughout the design and development processes necessary to reach that stage. It requires the application of a full, end to end, *systems engineering* approach to be successful.

The Turing award winning computer scientist Judea Pearl has recently noted three obstacles to the fulfillment of expectations for autonomous systems to exhibit “human level” intelligence [1] , which we consider to be similarly expressive of the attributes of a problem which indicate that its solution requires a *systems engineering* approach. The first is the requirement that the system be *adaptable* and *robust*, that it must be able to effectively operate under uncertain conditions and accomplish varying tasks. While there is always uncertainty in application, traditional engineering approaches have focused on eliminating uncertainty to largest possible extent by narrowing and restricting the channels available for the uncontrolled portions of the environment to impact the system (by insulating, compartmentalizing, modulating, etc.). These traditional approaches have seen their greatest and most rapid successes in cases where such narrowing and restriction is possible (development of integrated circuits, engines, chemical plants, etc.), but often the uncertainty and variability which a system is going to encounter cannot be removed by such processes, and must be accounted for by other means.

The second of Pearl’s requirements is that the system be *explainable*, so that it engenders user trust and allows engineers to diagnose and repair it in the case of errors. Much of the most recent and spectacular successes achieve in fields such as facial recognition or robotics have been driven by “black box” methods, with inner workings which are essentially opaque even to their designers. Establishing that such systems can operate safely and effectively becomes an empirical question, and addressing the system’s vulnerability to the potentially limitless number of possible adversarial or systematic errors becomes a significant burden. In contrast, *intelligible* and *interpretable* systems—employing “white box” or “grey box” models—allow for much more direct diagnostic approaches, as well as clear procedures for controlling for error. As an example, Caruana et al. [2] used a case where existing triage procedures introduced biases into clinical

outcomes related to pneumonia. These systematic biases obscured the underlying truth when models were fitted to the data, generating predictive algorithms that indicated that the presence of lung-diseases such as asthma actually lowered clinical risk for adverse events related to hospitalization for pneumonia! Their use of an intelligible, logistic regression based model—which allowed them to directly assess the probabilistic impact of all of the measured factors on the outcome by interpreting the coefficients in the linear predictor as changes in the predicted log-odds ratio—allowed for an accurate diagnosis of the cause. Patients with a history of lung diseases such as asthma who presented with symptoms of pneumonia were placed directly into intensive care, a procedure which more than offset the added risk incurred from the asthma. This reasonable and effective policy influenced the outcomes in such a way that an algorithm naively trained on the data, and developed in order to inform policy decisions, may have suggested that the effective policy be abandoned as counterproductive. Situations like this are especially present in the medical field, garnering the label of "confounding medical interventions" or CMIs [3]. The use of intelligible and interpretable models allows for biasing influences such as CMIs to be diagnosed and accounted for much more quickly than is possible when using black box models, allowing for system robustness and user trust to be more readily established within satisfactory bounds.

The third obstacle to general A.I. noted by Pearl is the necessity of integrating understanding of "cause-effect" relationships into the system. Of the three obstacles Pearl presents, this is the most difficult to translate into an equivalent systems engineering problem. But the essential task here is that the system (or the engineers presenting the system to users) need to be able to answer "What if?" questions: "What if I had done otherwise", "What if conditions x and y had not held?", etc. Such questions are the reason that system intelligibility and interpretability are desirable attributes, and answering them is necessary to construct robust and adaptable systems.

So addressing the first and second obstacles which Pearl presents requires something akin to answering cause-effect style questions, i.e. overcoming obstacle three.

In the research presented in this dissertation, we apply methods of statistical learning to the field of diabetes research, namely the development of predictive modeling algorithms for type-1 diabetes treatment. In doing so we encounter these three forms of problems/obstacles formulated by Pearl, and thus address the problem by using a systems engineering approach.

A more detailed discussion of the particularities of type-1 diabetes as a disease will be presented in the background in chapter 2 below, but here we note that it possesses many of the general features of medical conditions which present many challenges to engineers. Any engineering system designed to address medical conditions will have to take into account the ethical and practical constraints which pervade such problems. System users will vary, both in their physiology and behavior, and the system will need to operate in a way that accounts for the possible modes of failure or even misuse which are likely to be encountered in actual deployment. Thus *robustness* and *adaptability* are critical. Likewise, diagnosis of failures or errors, and user and physician trust throughout the development/deployment process are important components of any such system's success. Finally, the purpose of such systems will be to intervene with a specific patient, offering treatment advice or executing treatments themselves. So, as in the pneumonia example discussed above, it will be necessary to assess the impact of treatment policies on the data and deal with it appropriately.

The research which we will present in the following chapters develops a hypoglycemia (low blood sugar) forecasting system to be used in the context of a mobile health monitoring system for people with type 1 diabetes, which is applied to timeframes which exhibit a notable risk which is difficult to address by other means—during and after exercise and overnight.

At its basis are logistic regression models for each task which use sensor data available at the time of a system query to provide risk assessments and recommendations to system users in order to avert hypoglycemic events, without incurring undue exposure to elevated blood sugar (hyperglycemia). The choice of logistic regression is informed by the relative *intelligibility* and *interpretability* of such a model as compared to more sophisticated black box approaches, and its long use and familiarity to medical professionals and physicians whose agreement and trust in the system are critical. An additional boon to using logistic regression is its robustness to many kinds of unsystematic noise and its generalizability, due to its relatively simple linear structure of its classifications.

The core element of the research addresses the variability which such a system is going to encounter, both between different individual users and for each individual user his or herself over time, by means of an adaptive learning process which is called, for short, GMAdapt. This end-to-end online/transfer learning method is based on building a base, population level classifier and adapting it to an individual's unique data stream in order to quickly achieve a trackable, personalized model for making predictions. Like the logistic regression models which serve as its basis, a significant positive attribute of the GMAdapt procedure is its simplicity and explainability. We start with a population based model to ensure a reasonable prediction, then at each observation the model "steps" its defining coefficients in the direction indicated by the discrepancy between the observation and the current model's prediction.

Since the purpose of generating such predictions is to intervene in some manner in order to prevent adverse events, the questions of CMIs and how to handle them in such a system will need to be addressed. To do so, we will integrate a modification of the "net effect" simulation procedure [4], designed to assess alterations in insulin and meal delivery on blood glucose levels

in people with type 1 diabetes. Equipped with this simulation methodology we can simulate counterfactual scenarios to assess the impact of interventions on the outcomes and data, and compare it with alternative methods to do the same, such as deconfounding the data by controlling for interventions with dummy variables directly in the model. The final system will allow for rapid development of personalized hypoglycemia predictions, which is suitably robust to CMIs and testable in deployment in real world clinical trials.

The dissertation itself will be structured as follows. After the current introductory chapter, a second, background chapter giving the necessary preliminaries for understanding the following research will be presented. The next three chapters (chapters 3, 4, 5) will present in turn the basic GMAdapt procedure (chapter 3), the modified net effect resimulation method (chapter 4), and the integration of the two in an adaptive, personalized system for event driven hypoglycemia forecasting in the context of a mobile health decision support system (chapter 5). Each of these three core chapters will present the research following the format of a self-contained engineering journal paper, i.e. in the form of a content specific introduction, a methodological exposition, experimental assessment and evaluation either by simulation or retrospective real-world data analysis, and discussion of the outcomes of these assessments together with segues to the subsequent chapters.

The concluding chapter will summarize the main points and contributions of the research presented in fulfillment of the dissertation thesis statement and discuss the broader implications of this system, as well as the design processes and considerations which went into it, for other applications in other fields of engineering.

2. Background

2.1 Type 1 Diabetes

2.1.1 Pathology and Epidemiology

Type 1 diabetes (T1D) is an autoimmune disorder affecting the endocrine system which results in the destruction of the pancreas's beta cells. Consequentially, the body is unable to produce insulin, a hormone necessary for the effective regulation of glucose metabolism. While symptoms associated with the disease have been observed since antiquity [5], its exact etiology is still somewhat obscure. But both environmental and genetic factors are known to contribute to its inception and pathogenesis [6]. Symptoms usually first manifest during childhood or adolescence, with the majority of cases beginning before the age of 30 [7], however age is no longer defining factor for diagnosis [8], [9].

An elevated blood glucose (BG) level is the primary symptom of the untreated disease—clinically termed *hyperglycemia*. BG levels or associated factors have thus long served as the basis of diagnosis [10]. Without exogenous insulin injections, complications resulting from the production of ketone bodies produced by lipolysis in compensatory fat metabolism can lead to acute ketoacidosis, long term organ and tissue damage, and eventually death. However, exogenous insulin injections in T1D are not counter-regulated by the usual physiological feedback mechanisms which are present during the endogenous release of insulin in health. Thus excessive insulin injections can lead dangerously low blood glucose levels or *hypoglycemia* [11]. Hypoglycemia produces its own set of adverse symptoms, including acute cognitive impairment, seizure, coma, and death [12]. Achieving euglycemic control of BG by avoiding both hyperglycemia and hypoglycemia is the ultimate goal of new therapeutic technologies developed for the treatment of T1D. Of course, what constitutes excessive or appropriate insulinzation is

highly dependent on the individual patient and the context of the insulin administration. Sensitivity or resistance to insulin varies greatly from individual to individual, for a person themselves as a result of physical behaviors such as exercise [13] or physical activity [14], as well as due to cyclic hormonal changes related to circadian rhythms [15], or the menstrual cycle [16]. These factor need to be taken into account in order to achieve proper glycemic control and improve overall quality of life for people suffering from the disease.

Incidences of type 1 diabetes are increasing globally, with significant heterogeneity between ethnic groups and across geographical regions [17]. Overall estimates are that 1-in-300 persons manifest T1D by age 18 in the United States [18], and that there are approximately 1.3 million adults living with T1D in the United States as of 2016 [19]. The burden of the disease for individuals as well as public health systems is extensive, driving the development and application of new technologies for the treatment of the disease in order to improve medical outcomes and patient quality of life.

2.2 Current Digital Treatment Ecosystem for Type 1 Diabetes

Since the late 19th century physicians and physiologist hypothesized a relationship between the pancreas and glycosuria (the excretion of glucose in urine) [20], one of the earliest testable indicators of diabetes. This relationship was theorized to be mediated by some kind of secretion—termed “insuline"—derived from the “Islet of Langerhans” a region of the pancreas. Compounds resembling biological insulin itself as we know it today were first extracted in experiments performed by Banting and Best in 1920, leading to a Nobel prize in medicine for Banting and Macleod in 1923 [21]. Injections of this extracted insulin was able to alleviate the symptoms of T1D, eliminating glycosuria and reducing BG concentrations. Mass production of the extract was

soon undertaken by Eli Lilly using porcine pancreases, allowing the drug to be used in treatment for diabetics throughout the world [20].

Subcutaneous injections of this biological or “human insulin”, or its newer synthetic analogs developed since the 1990s [5], have served has as the primary medical intervention for controlling T1D from its discovery up until today. The conventional paradigm of insulin therapy for T1D involve syringe delivered basal and bolus injections to obtain euglycemic glucose control [22], commonly referred to as multiple daily injection or MDI therapy. *Basal* insulin injections are of long-lasting basal insulin meant to account for endogenous glucose production and physiological baseline metabolic needs. *Bolus* insulin injections are given to account for carbohydrates ingested with meals or elevated excursions in BG. Since the 1970s, continuous subcutaneous insulin infusion (CSII) pumps have provided an alternative means of insulin delivery to intermittent syringe injections and allow for basal insulin to be delivered continuously in a time varying profile which can account for variability in insulin resistance and sensitivity, mimicking action of the pancreas in health [23]. While patients themselves may prefer either of these options, meta-analysis indicate significantly better control outcomes while using less insulin overall for users of pump therapy [24]. Additionally, the continuous titration (in reality intermittent but frequent injections commonly termed “micro-boluses”) provided by CSII pumps lends itself more naturally to traditional control theoretic engineering practices.

Regular measurement of BG provides the feedback necessary to appropriately time and adjust insulin injections to achieve proper *euglycemic* control. As mentioned above, one of the earliest diagnostic assessments for T1D involved analysis of urine to detect glycosuria. Prior to 1965, when the first direct blood glucose testing strips using glucose oxidase was introduced, copper reagent based urinalysis was the only practical means of assessing BG. Advances in mobile

technology and blood assays allowed for the development of portable BG measurement technology and led to the introduction of the “self-monitored of blood glucose” (SMBG) treatment paradigm in the 1980s [25]. The introduction of SMBG allowed patients to regularly monitor their BG and adjust therapy accordingly throughout the day as well as keep records that could help inform their physicians of their BG variability in addition to the estimates available via blood assays of biomarkers such as HbA1c [26].

Starting in the 2000s continuous glucose monitoring (CGM) has been introduced, allowing for frequent sampling from subcutaneous tissue in order to achieved fine grained estimates of BG “24/7”, and more accurately assess overall and timeframe specific BG trends and variability [27]. Use of this technology has been expanding. The combination of CSII pump therapy together with CGM and the computing power and connectivity available on mobile smartphones has led to integrated treatment regimens such as “sensor augmented pump” (SAP) therapy [28], and produces obvious analogies with the kind of time-series dynamic systems commonly studied in the traditional electrical engineering areas of signal processing and control theory [29]. This suggests the possibility of creating “closed loop” feedback control style system in this application [30]. Devices for achieving closed loop control via intravenous insulin and glucose injections have existed since the 1970s, notably the BIOSTATOR [31]. Due to size and mobility constraints as well as other interfering factors, use of such devises was limited to in-patient studies and applications. The development of similar systems for obtaining closed loop control leveraging current mobile subcutaneous measurement and corresponding subcutaneous injection technology (s.c.-s.c) in recent years has been termed the Artificial Pancreas or AP project [32] [33].

The initial promise of the AP project was to provide fully automated “human-out-of-the-loop” control for people with T1D, similar to what is achieved by the functioning pancreas and

broader metabolic system in health. Despite recent advances, there is still significant room for other, avowedly "human-in-the-loop" approaches, to be designed and deployed while AP technology is being developed, and perhaps even indefinitely due to varying user desires for alternative treatment options [34]. These alternatives can take many forms, but the primary application for this dissertation will be decision support systems (DSSs) overlaid on to similar treatment technology ecosystems as present in SAP or AP systems i.e. with CGM, CSII pumps, and mobile smart phone connectivity.

2.3 Mathematical Modeling in Type 1 Diabetes

In addition to the aforementioned technological advances, concurrent conceptual advances have altered the way in which diabetes is understood, strongly influencing the design of treatments for T1D in recent years. By conceiving physiological variables as related components of system states, dynamic mathematical models allow for predictions of these state trajectories through time to be made based on known forcing inputs and initial conditions. Frequent measurements of glucose provided by CGM technology allow for the construction of time-series progressions of this primary variable of interest, which together with time-stamped records of meals and insulin injections lends the overall system itself to being represented by dynamic systems models. These dynamic systems models allow for simulations, predictions, and estimations of BG dynamics to be obtained in a transparent "white box" manner. Since integration of these sorts of theoretic models together with more empirical, data-driven approaches form one of the fundamental elements of the dissertation research presented in following chapters, we will briefly review some of these approaches, beginning with the foundational "Bergman minimal model" of glucose insulin dynamics.

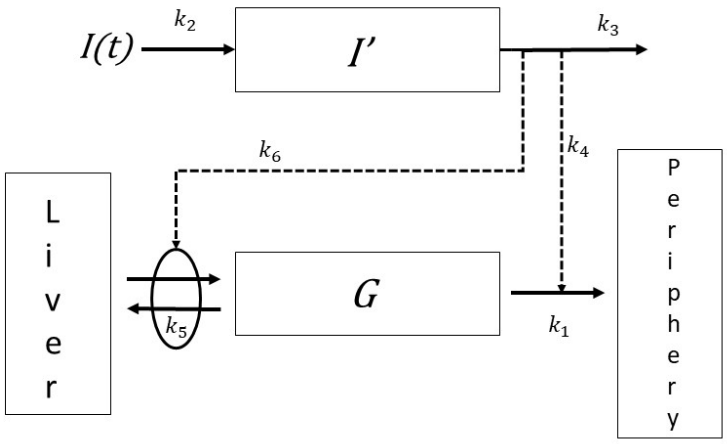
2.3.1 The Bergman Model

In 1979 Bergman et al. [35] proposed a series of seven compartmental dynamic response models of glucose-insulin metabolism—focusing on glucose kinetic disappearance. They evaluated each of the models for the identifiability, intelligibility, and number of parameters, as well as ability to fit data derived from glucose tolerance tests performed on laboratory canines. The chosen best model to aid in the task of insulin sensitivity estimation, model VI, can be represented by dynamic equations of the form

$$\frac{dG}{dt} = (p_1 - X(t)) \cdot G(t) + p_4 \tag{2.1}$$

$$\frac{dX}{dt} = p_2 X(t) + p_3 I(t) \tag{2.2}$$

This insulin dependent compartmental model of plasma glucose uptake in response to a direct plasma glucose injection was used by Bergman et. al. to assess insulin sensitivity. G represents plasma glucose mediating hepatic glucose balance, I plasma insulin appearance and X a compartment for “remote insulin” which facilitates glucose disappearance into periphery tissue.



$$X = (k_4 + k_5)I'$$

$$p_1 = -(k_1 + k_5)$$

$$p_2 = -k_3$$

$$p_3 = k_2(k_4 + k_6)$$

$$p_4 = B_0$$

$$p_5 = G_0$$

Figure 2.1: Compartmental diagram of the Bergman Minimal Model (VI) for estimating insulin sensitivity and parameter legend.

The nonlinear Bergman model has served as the foundation for further expeditions in the mathematical modeling of glucose-insulin dynamics based on common engineering methodologies for modeling closed-loop systems [36], and served as a catalyst in the design of other modeling and simulation platforms, including the UVa/Padova Type 1 Diabetes simulator.

2.3.2 The UVa/Padova Type 1 Diabetes Simulator

Experimental protocols for testing new therapies in T1D have traditionally required preclinical animal trials to ensure safety and effectiveness. The financial, time, and general resource burdens of conducting such trials are great and would significantly delay the deployment of complex and sophisticated algorithms such as required for AP systems. The constraints imposed by such

cumbersome trials provoked the development of simulation platforms based on dynamic models such as those proposed by Bergman decades earlier in order to allow for *in silico* preclinical trials bypassing the animal trial phase. These efforts culminated in the FDA acceptance of the UVa/Padova type 1 diabetes simulator to replace animal testing in preclinical trials for closed loop therapy [37]. The 300 total virtual subjects (100 adults, 100 children, 100 adolescents) in the simulator were represented by different parameter values of an underlying nonlinear compartmental glucose-insulin dynamic model, which were drawn from a joint distribution capable of reproducing the population level variability in BG dynamics observed in experimental studies [38]. These 26 free parameters for each *in silico* subject were chosen to represent the inter-subject variability in BG traces observed in clinical trials. Additionally, simulated models of CGM and CSII pump dynamics were integrated into the platform to facilitate testing on full *s.c.-s.c.* control systems such as mobile APs.

Over the years, new implementations and modifications of the UVa-Padova simulator have enabled *in silico* trials to be conducted which are better able to reproduce the kind of variability observed in the real-world T1D population: including more accurate insulin glucose dynamics and intraday variability allowing for multiple meal scenarios [39], and further refinements to the joint parameter distribution which generates the representative in silico population [40]. The simulator provides an excellent, transparent testing ground for evaluating potential new closed-loop therapies in T1D prior to conducting real-world clinical trials. However, it should be noted that the simulator as designed produces broad, population level assessments of the scenarios and therapies under examination— it does not as of yet provide an effective tool for providing individualized treatment tuning and optimization for real-world patients with T1D.

To provide assessments of potential new therapies at the patient level, methods of individualizing models or otherwise accounting for the variability observed in real-world signals in the T1D treatment ecosystem need to be develop. One such proposed method, based on CGM signal deconvolution with a linearized dynamic model of s.c-s.c. insulin glucose dynamics, allowing for time-frame specific reconstructions and simulations of insulin and meal regimes based on real-world data, the "net effect" simulator is presented in section 2.3.3 immediately below.

2.3.3 The Subcutaneous Oral Glucose Minimal Model and "Net Effect" Simulator

The UVa/Padova simulator and its associated *in silico* population allow for repeatable, population level experiments of different insulin therapies. While this is useful for broad assessments and for preliminary safety and feasibility studies and exploration, there are still often significant discrepancies between the consistent BG traces produced by its dynamic systems models and the variability in BG as assessed by observed CGM traces in real-world observations. To address these kinds of observed discrepancies, Patek et. al proposed a method of simulation integrating both theoretical compartmental models and empirical observations of CGM traces directly, the "net effect" (NE) simulator [4].

In its original form, the NE simulator is informed by actual records of insulin delivery obtainable from CSII pumps or recorded boluses as well as corresponding CGM data. The basic premise of the method is to use an extension of the Bergman minimal model called the SOGMM [41] or sub-cutaneous oral glucose minimal model to simulated the scenario recorded based on insulin records. The dynamics of the SOGMM model are represented by equations 2.3 and 2.4 below,

$$\dot{G}(t) = -\left(S_g + X(t)\right) \cdot G(t) + S_g \cdot G_b + (R_a(t))/V_g \tag{2.3}$$

$$X(t) = -p_2 \cdot X(t) + p_2 \cdot S_I(I(t) - I_b), \tag{2.4}$$

where $G(t)$ is the plasma glucose concentration (mg/dl), $I(t)$ the plasma insulin concentration (mU/L), $R_a(t)$ the plasma glucose rate of appearance (mg/kg/min), and $X(t)$ the proportion of insulin in the remote compartment (1/min). (For a more detailed explanation of the physiological meanings of these parameters and those that follow, as well as reference values, the reader is encouraged to consult [4], [42], [43].) On top of this "core" model of plasma glucose-insulin dynamic, the SOGMM includes subcutaneous insulin and gastrointestinal carbohydrate transport sub-models. Meal carbohydrate transfer through the gut is modeled sequentially via a two compartment sub-model,

$$\dot{Q}_1(t) = -k_\tau \cdot Q_1(t) + \omega(t) \tag{2.5}$$

$$\dot{Q}_2(t) = k_\tau . Q_1(t) - k_{abs} \cdot Q_2(t) \tag{2.6}$$

$$R_a(t) = \frac{k_{abs} \cdot f}{BW \cdot V_G} \cdot Q_2(t) \tag{2.7}$$

Where Q_1, Q_2 are the first and second "gut" compartments and $R_a(t)$ is the rate of appearance of glucose in the bloodstream, represented by the first state. Additional insulin kinetics, involving

subcutaneous injection (through a two compartment model) and measurement (as by a CGM) are given by the equations,

$$\dot{I}_{sc1}(t) = -k_d \cdot I_{sc1}(t) + u_i(t) \tag{2.8}$$

$$\dot{I}_{sc2}(t) = k_d \cdot I_{sc1}(t) - k_d \cdot I_{sc2}(t) \tag{2.9}$$

$$\dot{I}_p(t) = -k_{cl} \cdot I_p(t) + k_{c_1} \cdot I_{sc1}(t) + k_{c_2} \cdot I_{sc2}(t) \tag{2.10}$$

$$I(t) = (I_p(t)/(V_I \cdot BW) \tag{2.11}$$

$$\dot{G}_{cgm}(t) = -k_{sc}\left(G_{cgm}(t) - G(t)\right) \tag{2.12}$$

These nonlinear dynamics are meant to provide a simple representation of insulin glucose dynamics, and in order to implement this model in the net effect simulation procedure a linearization of the system (and further discretization) is used to create a linear time invariant (LTI) approximation suitable for the necessary computations.

The linearized system describes the evolution of the system by the equations,

$$\dot{x}(t) = Ax(t) + Bu(t) + G\omega(t) \tag{2.13}$$

$$y(t) = Cx(t) \tag{2.14}$$

where

$$A = \begin{bmatrix} -S_G & -G_b & 0 & 0 & 0 & 0 & 0 & \frac{k_{abs} \cdot f}{BW \cdot V_G} \\ 0 & -p_2 & 0 & 0 & \frac{p_2 \cdot S_I}{V_I \cdot BW} & 0 & 0 & 0 \\ 0 & 0 & -k_d & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_d & -k_d & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_d & -k_{cl} & 0 & 0 & 0 \\ k_{sc} & 0 & 0 & 0 & 0 & -k_{sc} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_\tau & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_\tau & -k_{abs} \end{bmatrix} \quad (2.15)$$

B = [0 0 1 0 0 0 0 0]', G = [0 0 0 0 0 0 1 0]', C = [0 0 0 0 0 1 0 0].

The eight states of the system, as ordered in the *x* state vector in the above equations, have the physiological interpretations given by

1. *Differential blood glucose concentration*
2. *Differential remote compartment insulin action*
3. *Differential interstitial insulin, first compartment*
4. *Differential interstitial insulin, second compartment*
5. *Differential plasma insulin*
6. *Differential interstitial glucose concentration*
7. *Differential gut, first compartment*
8. *Differential gut, second compartment,*

where the "differential" values are considered with respect to either reference or steady-state values. The time dependent insulin input, $u(t)$, calculated as a differential with respect to the basal profile enters into the first interstitial insulin compartment, while the net effect signal, $\omega(t)$ enters additively through the first "gut" compartment. A further discretization—using the "zero-th order hold"–assumption of this linear system allows for the deconvolution of the CGM signal with the

modelled dynamic to be calculated via a closed form solution, generating a net effect signal of the unaccounted for dynamics contributing to the observed variability. This signal can be "fed" back into the original SOGMM model along with known or modified inputs, providing a means of assessing counterfactual treatment scenarios within the specific time frame and for the specific subject from whose data the net effect was calculated.

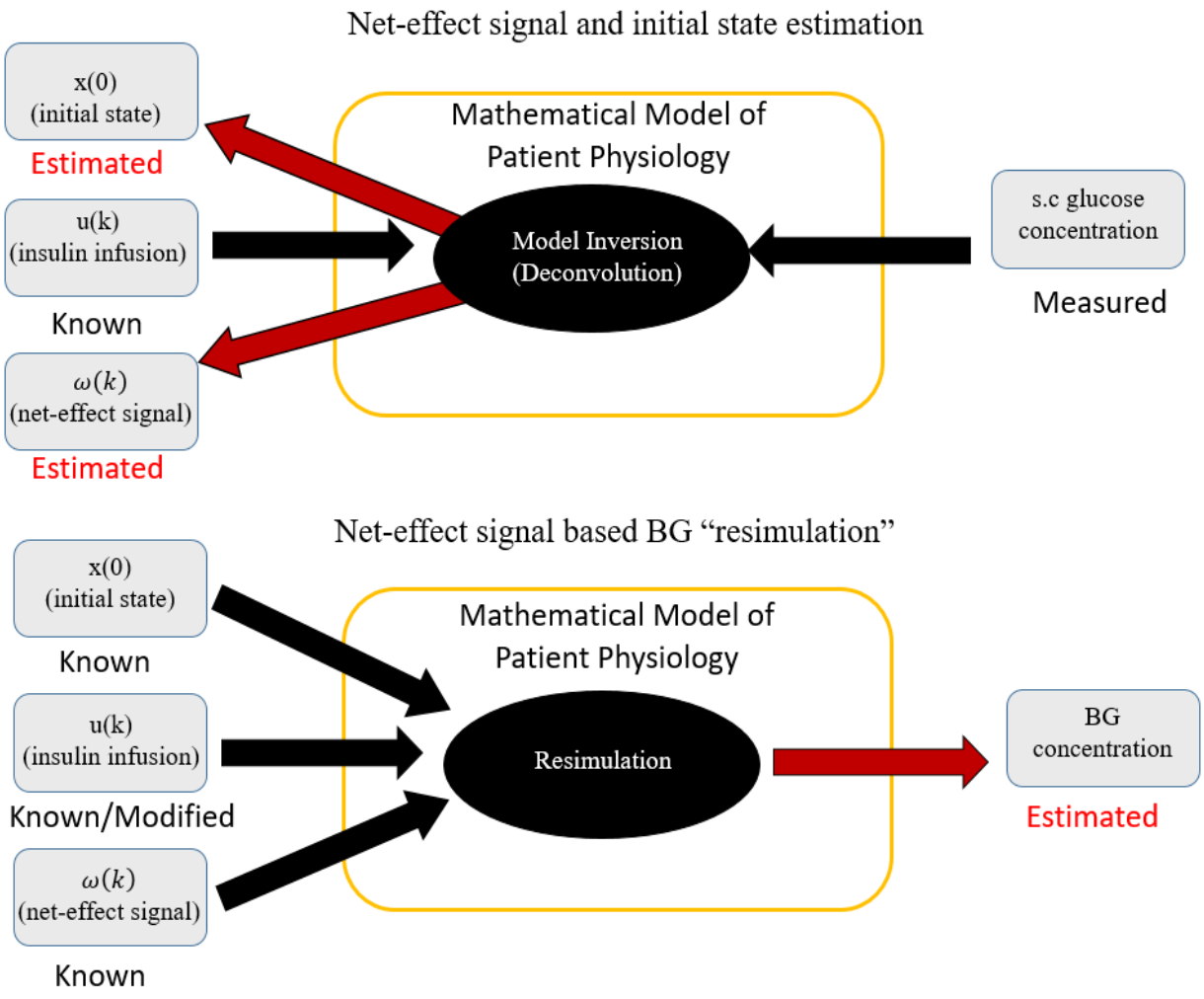


Figure 2.2: Diagram of the "net effect" simulation workflow

The crucial novelty introduced by the net effect simulator is that it allows for an individualized model of free-living observations to be constructed based on empirical records. This model takes into account sources of BG variation not explicitly modeled by the underlying compartmental/mathematical model by means of an additive disturbance term, and enables assessments of changes in insulin therapy specific to an individual and timeframe, *retro*dicting different outcomes as a result of changes in insulin titration policies and meal scenarios.

The net effect simulator has seen practical use, both in assessing the necessary accuracy need for non-adjunctive clinical use of CGM monitors (i.e. using them as the primary clinical means of assessing BG) [44], as well as assessing alterations of bolus therapy in response to changes in insulin sensitivity resulting from accumulated physical activity [45]. However, in the form presented in Patek et al. [4] there are several limitations which will be thoroughly addressed in chapter 4 of this dissertation. These include questions with regard to the extent that inter- and intra-subject variability in the underlying model parameters is absorbed into the net effect disturbance signal and verifying and assessing the domain of validity of the simulation procedure given the singular, unrepeatable nature of the simulated events [46].

This latter issue is perhaps the most important one from a conceptual or theoretical standpoint. The net effect simulator is a method which attempts to *retrodict* the BG dynamics generated by hypothetical or counterfactual insulin titration and meal scenarios with respect to singular, real world data. It in effect gives an answer to the question "*what would have happened on this unique day, for this unique person, under this specific scenario*". The subtleties of this statement shouldn't be lost. In the field of scientific model building, we seek to develop frameworks which allow us to predict what would *generally* happen in a given scenario under a certain set of circumstances. This is the basic conceptual framework assumed in both simulation

procedures such as the UVa/Padova simulator – where we use the in silico population as a stand-in for the general expectation of the response in the real world population of people with T1D – and in randomized control trials (RCTs)— where we wish to see if a certain therapy produces in aggregate sufficient positive outcomes to offset the various costs incurred. The subjects in the UVa/Padova simulator are fictions, and their individual responses are only of interest in assess possible bounds for worst or best case scenarios for a treatment in deployment. The design of RCTs is such that usually broad, population level outcomes are the focus of statistical analysis and clinical evaluation—again the effect of the treatment for a given individual participant in the study is only of interest with regard to establishing expectations of and bounds on worst case outcomes in cost-benefit calculations. The fundamental question for these methods to answer stated colloquially is *"what would happen for the relevant population generally, under this scenario"*.

In the terminology introduced by Judea Pearl [47], *in silico* studies conducted with the UVa/Padova simulator (or other similar simulation platforms) allow us to answer questions on "level 2" of the causal hierarchy i.e. questions dealing generally with the effect of *interventions*. The net effect simulator, on the other hand, is meant to answer questions posed on level 3 of the causal hierarchy, questions dealing with *counterfactuals*. Level 1 on the hierarchy are associational questions. Essentially, level 1 questions are those such as "Are taking aspirin and having a headache related events?", level 2 questions are those like "Does taking aspirin alleviate the symptoms of headaches?", while level 3 questions are those like "Did the aspirin *I* took after lunch cure *my* headache?"

While in recent years more and more of the social and medical sciences are adopting methods of causal reasoning and formulating questions on levels 2 and 3 of the hierarchy proposed by Pearl, many of the statistical procedures used in machine learning applications are—without

suitable modification— stuck on level 1. Many of these concepts will be relevant to the core of this dissertation research. But before then another critical element, “supervised learning” will be presented and discussed in detail in the immediate subsection below.

2.4 Supervised Learning in Type 1 Diabetes

The field of machine learning has traditionally been divided into two main subfields: supervised and unsupervised learning [48]. In unsupervised learning algorithms are used to extract patterns from mostly unstructured data. In the supervised learning setting, training data is structured so that examples of patterns of variables and explicitly labeled associated responses are given, and the task of the algorithm is to learn the relationship in the training data denoted by the labeled structure in a manner which is capable of being generalized to further, unobserved examples in the testing and deployment settings [49]. Usually, the tasks which methods of supervised learning address fall under either regression problems or classification problems. In the former, the goal of supervised learning algorithms is to determine a functional association between a structured set of feature variables (discrete or continuous) and a continuous response variable, in the latter, the goal is the same except for the response variable, which will be discrete and may represent categories of outcomes.

In T1D, the available data relevant to clinical tasks generally consists of more-or-less fixed demographic information regarding the individual with T1D, and, where/when available, data associated with the treatment ecosystem described in section 2.2. above. One of the first tasks that needs to be taken care of if we are to leverage this data together with newly developed techniques and advances in machine learning is to properly formulate problems so that the technology and knowledge we have can effectively be brought to bear at the clinical tasks which need to be

accomplished to help alleviate the symptoms of the disease and improve quality of life for patients. One of the most straightforward proposals is to formulate the task of predicting future BG levels—the direct, clinical outcome of interest in T1D treatment— based on available data as a supervised learning problem.

At the highest, most abstract level this is a rather simple problem to state: given the historical data, the records available, and the outcomes observed, we must generate a model capable of predicting adverse BG events (hypoglycemia or hyperglycemia) far enough ahead of time so that those events can be prevented. However, this problem is more difficult to solve in practice than to state in theory. Compartmental, dynamic models such as the Bergman model, SOGMM, and the nonlinear model undergirding the UVa/Padova simulator discussed above are built around a large accumulation of clinical and engineering experience, and lend themselves to broader assessments of the expected outcomes of therapies and treatment regimes. But this task differs from that of data-driven, short-to-medium range predictions of BG for individuals in free-living conditions, which would allow for near-term corrective actions to be taken. For that task, alternative methods need to be developed.

Perhaps the most direct approach to incorporate the data-streams available in an s.c-s.c treatment ecosystem is the fully closed-loop A.P systems such as briefly touched on in section 2.2 above. Much like the field of “self-driving cars”, the progress of the technology is such that there a continuous advance of automation, slowly progressing, rather than the discrete jumps that may be imagined by the popular media. Proprietary predictive algorithms integrated with newer CGMs which modulate subcutaneous insulin infusion in response to predicted hypoglycemia, such as the Tandem Diabetes Care “Basel-IQ” system, are already being evaluated in large scale studies [50] and are seeking FDA approval for real-world clinical implementation. But there are limitations to

such approaches. Namely the usable prediction horizon when directly modeling BG concentration based on CGM and pump data by current methodologies is 30-60 minutes [51]. For systems with regular actuation, such as APs or other control algorithms that modulate insulin infusion continuously or regularly via a CSII pump, and in situations without predictable, abrupt negative disturbances, this does not pose too much of an issue. However, for alternative approaches to using technology to improve treatment of T1D, such as DSSs, which lack continuous or regular means of effective actuation of the system, or for significant short-term, difficult to predict disturbances such as bouts of exercise, these limitations may prove cumbersome and prevent effective system implementation.

One method of dealing with the current limited effective prediction horizons obtainable by methods of direct BG forecasting is to reframe the task as a classification problem. The tripartite clinical division of BG levels into hypoglycemic, euglycemic, and hyperglycemia ranges lends itself naturally to a categorical mapping. Since current accepted practice places no particular weight on differing values of BG within the euglycemic range (70-180mg/dl), it is natural to conceive of exiting this range—either into hyper or hypoglycemia— as a discrete event, amiable to a binary classification approach. In the core of the dissertation research presented in chapters 3-5 below we will use this later approach to develop personalized hypoglycemia prediction algorithms capable of implementation in the context of DSSs via adaptive generalized linear models. But prior to that some background on the general process of regression and classification modeling which will be used is presented in the subsections below.

2.4.1 Elementary Linear Modeling

The most well-trodden task of supervised learning for regression is the elementary statistical problem of generating a linear model based on give set of *N* observations of responses, each associated with *k* feature covariates, $\{x_i,\ y_i\}_{i=1}^{N}$. In this case, the usual assumption is that response observations, i.e. $y_i \in \mathbb{R}$ are generated by linear combinations of the feature variables in the vector $x_i = \left[x_{i,1},\ x_{i,2}, \dots, x_{i,N}\right]$ and an additive zero-mean Gaussian distributed noise term, ϵ_i so that

$$y_i = x_i\beta + \epsilon_i, \tag{2.16}$$

where $\beta = [\beta_1,\ \beta_2, \dots, \beta_N]$ is a vector of the regression coefficients. Given a set of N observations, $\{x_i,\ y_i\}_{i=1}^{N}$ fitting the Gaussian noise assumptions, equation (2.16) can be extended to the matrix form

$$Y = X\beta + \boldsymbol{\epsilon}, \tag{2.17}$$

where $Y = [y_1 \mid y_2 \mid \dots \mid y_N]^T$, $X = [x_1^T \mid x_2^T \mid \dots \mid x_N^T]^T$, and $\boldsymbol{\epsilon} = [\epsilon_1 \mid \epsilon_2 \mid \dots \mid \epsilon_N]^T$. In this case, the least-squares fit provides an unbiased estimate of the β coefficients and is given by the closed form solution

$$\widehat{\beta} = (X^T X)^{-1} X^T Y. \tag{2.18}$$

Such basic linear models and extensions of this procedure have seen some application to problems of glucose prediction in T1D, especially in the form of time-series regressions [52], [53]. However, for the purposes of applications such as prediction of hypoglycemia, the limited empirical prediction horizon (approx. 20 -30 minutes) obtainable by such models means that

alternative and perhaps more sophisticated approaches such as generalized linear models are need to enable extended prediction horizons.

2.4.2 Generalized Linear Models

Generalized linear models (GLMs) serve as an extension of elementary linear models by allowing for nonlinear relationships between the response and feature variables by means of a canonical *link* function [54]. In the case of the standard or elementary linear model the relationship given (2.17) can be thought of in terms of mathematical expectation, and formulated as

$$\mathrm{E}[y_i] = \mu = x_i\beta \tag{2.19}$$

using the linearity of the expectation operator and the zero mean Gaussian assumption regarding the distribution of the ϵ_i. In the elementary linear modeling case, the link function is simply the identity function, $g(x) = x$. But more elaborate links are possible, and the use of such link functions allow for different kinds of models of the relationship between the feature and response variables to be constructed.

The most important family of such link functions for the purposes of this dissertation is the exponential family of distributions, having the form [55]

$$f(y \mid \theta, \varphi) = exp(a(\varphi)\{y\theta - g(\theta) + h(y)\} + b(y, \varphi)\ ). \tag{2.20}$$

For cases of binary classification (such as the prediction whether or not a hypoglycemia will occur given the observed data) we want to develop a GLM capable of representing Bernoulli random variables. Letting π be the mean expected outcome a single Bernoulli variable will be distributed,

$$f(y \mid \pi) = \pi^{y}(1-\pi)^{1-y}$$

$$= \exp(\log(\pi^{y}) + \log((1-\pi)^{1-y}))$$

$$= \exp(\mathrm{y} \cdot \log(\pi) + (1-\mathrm{y}) \cdot \log(1-\pi))$$

$$= \exp(y \cdot \log\frac{\pi}{1-\pi} + \log(1-\pi)), \tag{2.21}$$

which we note is in the form of an exponential family distribution with $\theta = \log\frac{\pi}{1-\pi}$, $g(\theta) = \log(1-\frac{e^{\theta}}{1-e^{\theta}})$, $h(y) = 0$, and $a(\varphi) = 1, b(y,\varphi) = 0$ (suppressing the dispersion parameter φ, which is not necessary in the Bernoulli or binomial distributions). This forms the basis of a specific instance of a GLM, namely the logistic regression for the hypoglycemia forecasting algorithm composing the core of the dissertation research.

2.4.3 Logistic Regression

A logistic regression is perhaps the most straightforward method of generating a discrete, categorical classifier on the basis of an underlying generalized linear model and data composed of observations of associated feature variables and discrete binary outcomes, which appropriately models the outcome by means of probability. The basic assumption is that the outcome variables each follow a Bernoulli distribution conditional on the feature variables, i.e.

$$E[y_i|x_i] = \pi_i, \tag{2.22}$$

where π_i is a function of x_i and a fixed set of parameters, β. This distribution is assumed to be binary so that $y_i \in \{0,1\}$, and can be written in the form

$$f(y_i \mid \pi_i) = \pi_i{}^{y_i}(1-\pi_i)^{1-y_i} \tag{2.23}$$

To relate the data x_i to the expected value π_i the logit link function is used, so that

$$\log \frac{\pi_i}{1-\pi_i} = x_i\beta. \tag{2.24}$$

This leads naturally to the sigmoidal representation of π_i,

$$\pi_i = \sigma(x_i\beta) = \frac{1}{1+e^{-x_i\beta}} \tag{2.25}$$

Whereas Gaussian linear models allowed for direct closed-form solutions to determine unbiased estimates of the β parameters, no such solution exists even for the relatively simple case of a logistic regression GLM. Instead, the usual approach to estimating the β coefficient parameters is via maximum likelihood estimation (or MLE) procedures. The basis of MLE is to find the mode of the likelihood function (the probability density considered as a function of the parameters given observed resolutions of the random variables). For the case of logistic regression on N independent observations of data, the likelihood function will be given by

$$L(\beta|x_i, y_i) = \prod_i^N \sigma(x_i\beta)^{y_i}\big(1-\sigma(x_i\beta)\big)^{1-y_i} \tag{2.26}$$

For mathematical convenience the monotonic log transform of the likelihood is used for MLE estimation, which amounts to maximizing the function

$$l(\beta|x_i, y_i) = \sum_{i=1}^n y_i \cdot \log \sigma(x_i\beta) + (1-y_i) \cdot \log\big(1-\sigma(x_i\beta)\big) \tag{2.27}$$

also known as the *cross-entropy* loss.

The most common methods for maximizing the cross-entropy loss of logistic regression are based on extensions and modifications of Newton-Raphson maximization/minimization procedures [56]. Several such methods will be described in detail in the subsection dealing with optimization and model fitting below (section 2.5).

2.4.4 Accounting for Heterogeneity--Mixed and Hierarchical GLMs

The structure of equation (2.22) allows for the construction of linear regression models which generate outcomes which can be interpreted as probabilities [57]. This model can be extended to include "noise" beyond what is present due to the nature of the binary outcome generated by the π_i terms. One common way is to assume the linear predictor component is itself randomly distributed according to a conditional Gaussian random variable, i.e.

$$\log \frac{\pi_i}{1-\pi_i} = x_i\beta + \sigma_i, \tag{2.28}$$

where the σ_i are independently identically distributed (i.i.d) Gaussian random variables with mean 0 and standard deviation σ. The assumption here is that any given observation follows this same distribution, and that the β which define the model are *fixed* parameters. The relationship between outcomes and feature variables is assumed to be the same for every observation in the dataset, excepting the additive noise terms σ_i.

It is often the case, however, that there is heterogeneity with regard to the relationship between feature variables and outcomes, such as the case when samples are drawn from an

aggregated population consisting of distinct, individual subjects or subgroups. In the context of logistic regression, when information about such subgroups is explicitly available, this heterogeneity is commonly modeled by assuming that for the *i-th* observation from the *j-th* subgroup the log odds ration of the probability has the form

$$\log \frac{\pi_{i,j}}{1-\pi_{i,j}} = x_{i,j}\beta + z_{i,j}b_j + \sigma_{i,j} \,. \tag{2.29}$$

This is referred to as a generalized linear *mixed* models (GLMM) due to the addition of *random* effects (the b_j vectors) to the fixed β vectors in the model structure (which can also be viewed as a mixture distribution) [58]. Much of the literature on the subject of GLMMs focuses on the effects of the "randomness" of the subgroup specific b_j terms on the overall estimation and inference process with regard to the fixed β coefficients [59]. Though there is much ambiguity both conceptually and practically in the separation of so-called fixed and random effects, we will stick to the definition in terms of the fitting procedure— definition 5 in Gelman's taxonomy [60]—to avoid more qualitative and subjective considerations in regards to this ambiguity. Much like the noise introduced by the σ_i terms, the b_j terms are often treated as a nuisance variables arising due to the correlation between within subgroup observations, which to lead to *overdispersion* of the empirical variance and thus impacts statistical inferences made based on the model [61]. This is especially the case when the population level fixed effects are considered the primary or sole effect of interest. However, the information within b_j values is of great importance if the level of intervention or treatment for which the model is going to be used is that of the subgroup or individual subject themselves.

In the case of developing personalized predictive hypoglycemia forecasting models for people with T1D, the point of interest will be the individual patients. In particular, we will be interested in construction models with "varying intercepts and slopes" [62], which allow for the relationships between feature variables and outcomes to vary across the individual subjects. In our case we will be interested in the values of the specific, individual parameters—the $b_j's$—since our goal will be to develop *personalized* predictive models for hypoglycemia forecasting. The general problem of determining a specific subgroup's or subject's b_j, as opposed to the focus on zero mean assumption and variance that follows when such parameters are considered a nuisance to the estimation of the population level $\beta's$ and subsequent inferences, is called a random effects *prediction*, as opposed to a fixed effects *estimation*, problem [63]. In the case of GLMMs, these predictions are generally obtained via Bayesian methods [64]. For instance, the Gibbs sampling Markov Chain Monte Carlo (MCMC) approach proposed in [65] assumes that the b_j follow a prior distribution with mean zero and covariance matrix (itself to be estimated) D, and that the outcome data y_i follow a conditional exponential family distribution $f(y_i|\beta, b_j, D)$, and thus we have

$$f(b_j|y_i) = \frac{\int\int f(y_i|\beta,b_j)g(b_j|D)p(\beta,D)d\beta dD}{\int\int\int f(y_i|\beta,b_j)g(b_j|D)p(\beta,D)db_j d\beta dD} \tag{2.30}$$

where $g(b_j|D)$ is the conditional Gaussian distribution density, and $p(\beta, D)$ the joint prior density of the β, D parameters. Direct numerical evaluation of this integral (e.g. in order to assess the expected value of b_j given the data, $E[b_j|y_i]$) is intractably difficult to accomplish for many applications, necessitating relatively computational intensive MCMC approaches [66].

Additionally, the extensive computation burden grows with the number of variables, observations, and subgroups present in the data.

2.4.5 Evaluation of Clinical Classifier Performance

Evaluation of modelling performance is task and context dependent. For the case of continuous prediction models such as linear regression, the most well-known and commonly used metrics to assess performance are the mean-square error (MSE) or root mean square error (RMSE). For the case of classification models perhaps the most straightforward metrics are the expected or empirical risks obtained via assuming a "zero/one loss" function [67] [68]. For example, if $f(x,\beta)$ generates a predicted value in $\{0,1\}$ (binary classification) and $y \in \{0,1\}$ represents the actual value associated with the observed x data, then a loss function can be constructed such that

$$L(x,y,\beta) = \begin{cases} 1, & f(x,\beta) \neq y \\ 0, & f(x,\beta) = y. \end{cases} \tag{2.31}$$

If x and y are generated by a random data generating process (either due to sampling by the experimenter or nature itself) following a joint cumulative distribution $P(x,y)$, then the expected risk for a given set of parameters β is given by

$$R(\beta) = \int L(x,y,\beta)dP(x,y). \tag{2.32}$$

And the empirical risk for a given data set of N observations is given by

$$R_{emp}(\beta) = \frac{1}{N}\sum_{i=1}^{N} L(x_i,y_i,\beta). \tag{2.33}$$

For fixed data sets, model fitting procedures such as discussed below in section 2.5 generally focus on minimizing this empirical risk given in equation 2.33. In the case of classification problems, this amounts to maximizing accuracy or the number of correct labels assigned by the classification function $f(x, \beta)$. However, this metric can be misleading in the clinical setting, where all "misses" are not equal and the costs associated with false-negative and false-positives may diverge greatly. Also, the events themselves may be rare, leading to unbalanced data, so that accuracy levels which prima facie appear impressive (say 99%), may in actuality not represent a significant gain over uninformative classification methods (such as "always label 0").

In the clinical setting, methods of evaluating classifiers beyond the accuracy metric which incorporate information about the trade-off between different kinds of mislabeling error (false positive and false negatives) are preferred, such as the receiver-operator characteristic or ROC curve [69] [70]. The ROC curve presents graphically the relationship between false-positive and false-negative rates as difference classification thresholds are chosen.

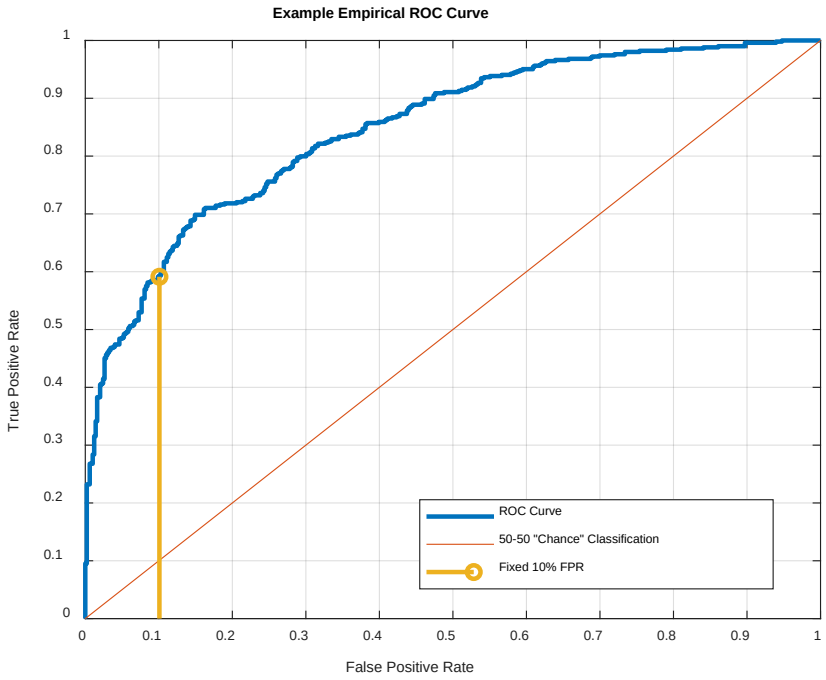


Figure 2.3: And example of an "empirical" ROC curve, representing the trade-off between true-positives and false-negatives on a simulated data set. Red line indicates "chance" performance obtained by a "coin flip" appropriately weighted for the class prevalence (50-50 for balanced data). The yellow stem plot gives the performance at a fixed 10% false positive rate.

In addition to the graphical representation offered by the plot of the ROC curve, the area under the roc curve (ROC-AUC) or "c-statistic" serves as a common single number measure of classification performance in the context of clinical applications [71]. The graphical interpretation of the ROC-AUC is straightforward, but in classification terms it also is associated with the probability that a given pair of sampled observations with one member from each class will be appropriately labeled by the classifier [72]. While commonly helpful in representing overall classification performance, the single value summary provided by the ROC-AUC may not be appropriate for a given clinical application. For instance, treatment constraints may impose

conditions such as no more than ten percent false positive rate is acceptable in practice, and the nature of the ROC curve allows for degenerate cases where performance is exceptional in certain regions, but poor in the area of interest—a case that cannot be directly ascertained from the AUC alone. While visual inspection of the curve can provide insight in these cases, fixed statistic such as partial AUCs (pAUCs) or fixed false positive performance can help evaluate relevant performance when such constraints are active.

Before leaving this section we note that the general interpretation of the empirical ROC curve, that the true positive or detection performance on the y axis is obtainable with false positives not exceeding the associated value on the x axis for a given threshold, is dependent on a fixed data setting. In more sophisticated settings such as we will deal with in the core chapters 3-5 below, interventions will be performed at varying rates *depending on the chosen classification threshold,* and these interventions will impact further classification performance. The issues involved in online-intervention systems such as will be presented require subtler insights than naïve ROC analysis can provide. A new method of data visualization, presenting the trade-off between overall or specific classification performance as represented by the ROC-AUC or some other suitable single value metric and the aggressiveness of the intervention policy will be presented in chapter 5 below.

2.5 Optimization and Model Fitting

In equation (2.18) in section 2.4.1 we gave the closed form solution for the estimation problem with regard to data under the assumption of a linear, Gaussian distribution of basic linear modeling. No such closed form solution exists for generalized linear models such as those used in logistic

regression. Instead, numerical approximation techniques are used in order to achieve appropriate estimates of the β coefficients.

As mentioned above, the workhorse numerical technique for this and similar problems is maximum likelihood estimation (MLE). In the MLE of logistic regression, we seek to conveniently maximize the likelihood function given in equation (2.26) by performing the equivalent minimization of the negative log likelihood given by the negation of equation (2.27) with respect to the training data available. Many numerical procedures are available for accomplishing this task (e.g. [56]) but the most commonly used for statistical computation packages are variations on Newton-Raphson methods which iteratively estimate β parameters according to the mapping

$$\beta^{(n+1)} \leftarrow \beta^{(n)} - \boldsymbol{H^{-1}} \nabla l\left(\beta^{(n)}\right), \tag{2.34}$$

where $\boldsymbol{H^{-1}}$ is the inverse of the Hessian matrix (or some computationally suitable approximation) associated with the log-likelihood function $l\left(\beta^{(n)}\right)$, and $\nabla l\left(\beta^{(n)}\right)$ is its associated gradient vector of the function with respect to β, evaluated at the iterate $\beta^{(n)}$.

The convergence of the iterative procedure defined in (2.34) to a global minimizing function of the empirical risk for a given set of training data and convex loss function such as (2.27) when the data are not separable [73] is well established, and separable cases can be handled by modifying the likelihood with prior information and moving towards a maximum a posteriori (MAP) estimate (equivalently conceived of as a adding a regularization term to the likelihood cost function) [74].

In modeling situations where the Hessian matrix is not easily ascertainable, such as Artificial Neural Networks, alternative approaches using only first-order moment information (i.e.

only using the $\nabla l(\beta^{(n)})$ term) such as "gradient ascent/descent" algorithms, have achieved great popularity due to their feasibility in such scenarios and much lower computational burdens imposed [75]. A gradient descent algorithm (for loss function minimization) has the iterative form

$$\beta^{(n+1)} \leftarrow \beta^{(n)} - \gamma_n \nabla l(\beta^{(n)}), \tag{2.35}$$

where the $\gamma_n \in \mathcal{R}$ are a chosen sequence of positive real numbers. When suitable chosen and under convex regularization and functional assumptions, the γ_n ensure convergence to a global minimizer of the empirical risk of the $l(\beta)$ functional on the training data.

In addition to the gradient descent calculated with full training data as in (2.35), there are also methods of *batch* or *stochastic* gradient descent (SGD) where the updates are calculated with respect to a sampled subset or individual observation drawn from the training data [76]. The latter, stochastic gradient descent, can also be readily implemented *online* as examples are drawn directly from the distribution of the data generating process (DGP) itself, as opposed to sampled from a fixed training data set [77]. Conceptually, this changes the problem from the minimization of the empirical risk on the training data, to the minimization of the expected risk generally, based on the pseudo-infinite data stream drawn from the DGP. In practice, it is straightforward in that instead of the full gradient being calculated on the training set, the gradient is calculated at each iteration only with respect to the current observation $\nabla l(\beta^{(n)}, x_n)$, which under the assumptions of independence, provides an unbiased estimate of the true gradient

$$\beta^{(n+1)} \leftarrow \beta^{(n)} - \gamma_n \nabla l(\beta^{(n)}, x_n). \tag{2.36}$$

The foundation of the dissertation research is to leverage the structure of the event-driven hypoglycemia prediction problem in T1D together with online optimization methodologies like SGD to produce personalized predictive models within relevant clinical and time constraints.

2.6 Personalized Predictive Models for Hypoglycemia Forecasting

The purpose of the foregoing background is to establish the basis for presenting the core contribution of the dissertation research that follows— building personalized predictive models under the usual constraints of the T1D treatment ecosystem.

The reasoning motivating the goal of developing predictive models is clear enough—by predicting adverse events ahead of time it may be possible to prevent those events and the corresponding complications by taking precautionary action, without unnecessarily exposing oneself to the tradeoff effects that such actions may entail. For instance, in the context of T1D, future hypoglycemic events may be eliminated by ingesting unbolused carbohydrates or hypotreatments or by a reduction in basal insulin delivery for CS-II pump users. But both of these actions raise the possibility of greater exposure to hyperglycemia—especially when the future trajectory was not liable to go low in the first place and the carbohydrate ingestion was unnecessary.

More interesting, and requiring more explanation, is the need and desire to develop *personalized* models, tuned to individual patients. There has been a recent push to develop personalize or precision medicine generally [78], and for diabetes in particular [79]. To explain this phenomena, we should point to one of the limitations of what is considered the gold standard for clinical experiments the randomized control trial (RCT). As discussed in section 2.3.3, RCTs

are primarily meant to establish the causal, broad population level effect of a treatment intervention. However, in the presence of significant heterogeneity in treatment responses, the randomization which allows for confidence to be placed in the causal effect demonstrated by the study overall may mask more interesting and varying phenomena occurring at the point of treatment. In the most extreme case this can result in an instance of "Simpson's Paradox" [80], [81], where aggregated data displays trends and generates inferences that are inverse of what is displayed by each of the subgroups which taken together constitute the data.

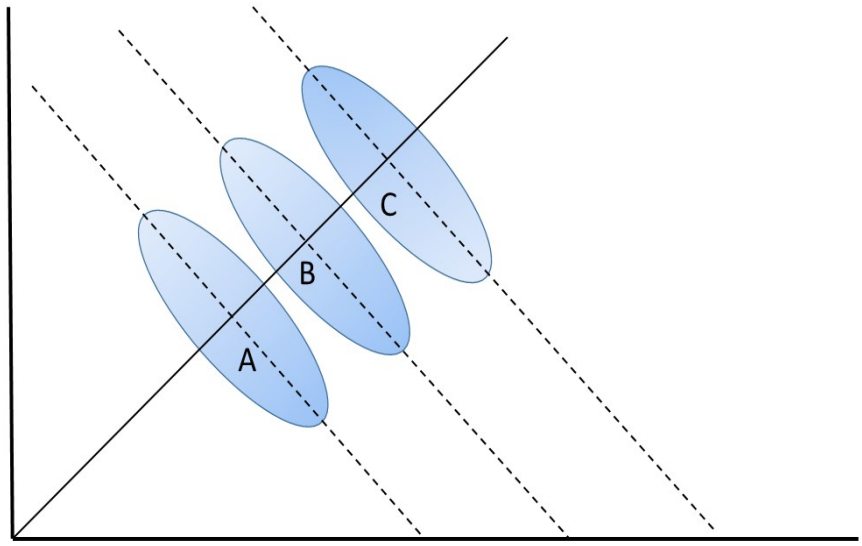


Figure 2.4: A qualitative demonstration of Simpson's Paradox. In this case, each of the labelled subgroup distributions: A, B, C, is oriented so that a negative trend in the relationship between the features represented by the horizontal and vertical axes is present. In aggregate, however, a positive trend is present.

The reversal of trend in full cases of Simpson's paradox as presented graphically in fig. 2.4 is an extreme example. But milder cases of discrepancy in relationships between predictor and

response variables between subgroups (such as individual patients for the case of medical treatments and interventions) is likely to occur due to inter-subject physiological heterogeneity. The use of personalized or precision medicine is a means of avoiding the problems induced by using “one size fits all” approaches on heterogeneous patient populations by leveraging either “omic” data (genomics, epigenetics, proteomics, metabolomics, etc. [82]), or electronic health records (EHRs) featuring medical history and demographic information relevant to treatments [83].

The continuous monitoring of BG made possible by CGMs as well as insulinization records tracked by CS-II pumps present unique opportunities for using the surfeit of data generated to obtain personalized treatment models for people with T1D. Before laying out the basic setup that we propose for achieving such personalization in the context of hypoglycemia prediction, we will discuss the normal level of personalization achieved by standard insulin therapy and the time-series setting in which such algorithms will be developed.

2.6.1 Current Personalization of Insulin Therapy

The current paradigms of “open loop” CS-II pump therapy and MDI basal therapy each already incorporate a certain degree of personalization. In the case of open-loop CS-II pump therapy, three parameters which allow for personalization of insulin therapy are the basal insulin profile [84], the carbohydrate-to-insulin ratio, and the correction factor [85], these latter two factors are paired with injections of basal insulin in the case of MDI therapy [86]. In common practice, each of these component parameters of treatment are personalized for the individual by their physician based on demographic factors and clinical/empirical outcomes. Recently there have been attempts to

personalized therapy via automatic, data driven approaches and heuristics for accomplishing this task [87].

The purpose of the carbohydrate-to-insulin ratio or CR factor is to enable appropriate bolusing of insulin to account for meal carbohydrates, while the correction factor is meant to calculate appropriate doses to lower elevations in BG to target levels. The basal profile is set on an insulin infusion pump in order to account for endogenous glucose production and insulin needs due to the regular course of metabolism and varying insulin sensitivity, independent of meals. This delivery generally follows a circadian pattern of insulin infusion varying throughout the day. Likewise, the CR and CF can vary over the day to account for changes in insulin sensitivity and thus can also be associated with time-varying profiles.

2.6.2 Leveraging Data from T1D Treatment for Model Personalization

The data available in the mobile T1D treatment ecosystem comes primarily in the form of time series. For instance, a CGM produces a time series of glucose measurements that can be digitally recorded and collated with similar digital insulin infusion records from CS-II pumps or “smart-pens”, usually partitioned into five minute intervals, with 288 readings per day. Likewise, estimated meal carbohydrates—either associated with insulin boluses logged by the pump or recorded by the user themselves in a digital dairy— can also be readily associated with corresponding five-minute time windows, and these data can be organized as associated time series threads.

The approach we intend to use in the research presented in the chapters below is one of predictive forecasting using these time series data both as inputs and to determine responses [88] formulated as a classification problem modeled by a logistic regression. To arrange this data in a

manner amiable to a logistic regression based forecasting algorithms, we need to identify feasible windows both for the extraction of predictor variables and for recording the resolution of the observation event labels. In general, these windows and the relevant variables will be task dependent. In our main applications, we wish separately to predict potential hypoglycemia related to exercise and hypoglycemia occurring overnight. In each of these cases, current short term prediction and forecasting methods based on pump, meal, and CGM data face significant challenges. In the former case, the known “future” disturbance of impending exercise will not be ascertainable from the past data alone and will have a major impact on BG dynamics. In the latter case, sleep will interfere with the ability of the user to respond quickly to changes, and the extensive overnight period will require longer range forecasts than available by means of short-term extrapolation based predictors. Thus, both of these cases present opportunities for the application of DSS style systems that leverage a classification based approach to develop longer term and known future disturbance tuned forecasts.

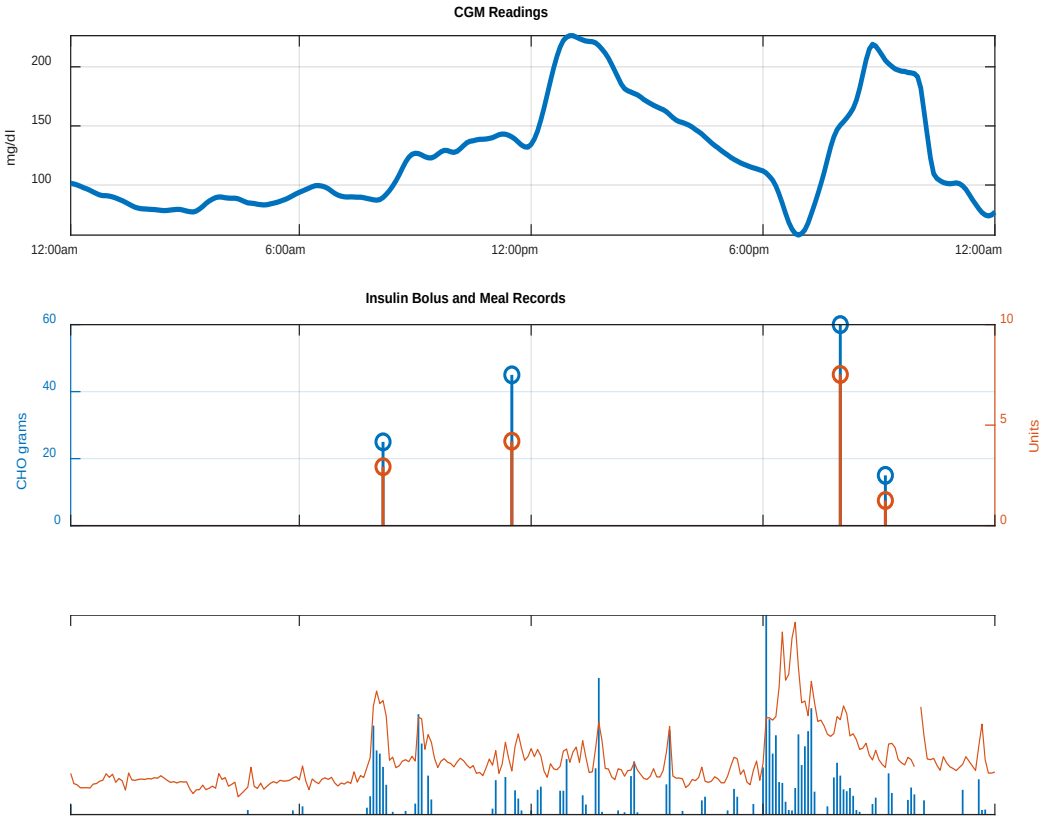


Figure 2.5: Visual representation of time series data available in T1D treatment ecosystem. CGM, insulin bolus, meals, as well as heartrate and step count records from wearable trackers are aligned by time and can be used to adjust or inform treatment decisions.

To formulate the time series data in a manner which allows for GLM based classifiers such as logistic regression to be used for hypoglycemia prediction and prevention systems, query points need to be established to orient the system and to inform the user or their DSS in an intelligible and actionable manner. These queries can be either triggered manually by the user—such as via a button push prior to exercise or bedtime—or "event triggered" at associated times or as a result of specific attributes of the data (e.g. BG readings or rates of change obtaining certain thresholds).

Since the purpose in our application is forecasting/prediction, it is necessary that the predictor variables be derived from data available at the system query time, and thus the predictor windows will need to be generated so that they are associated with times prior to the given query. The resolution window—where the observations class label is determined—must cover some time frame after the query and the predictor data window.

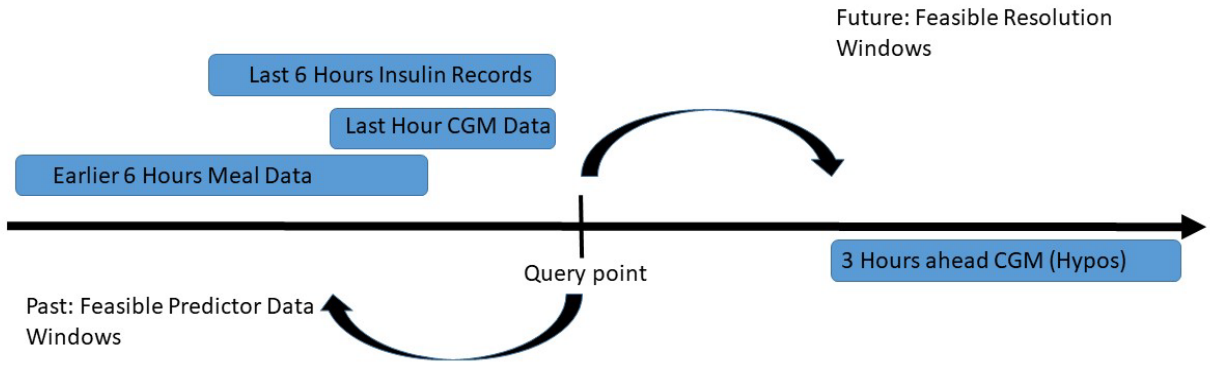


Figure 2.6: Demonstration of time-series structure of prediction and resolution windows for hypoglycemia forecasting.

Once a suitable set of feature variables and the associated responses has been extracted from the time-series data, standard fitting methods such as those mentioned in section 2.5 or off-the-shelf software packages for fitting generalized linear models can be used to generate logistic regression based classifiers just as in the case of static, non-time series data.

Chapter 3

3.1 INTRODUCTION

In the current chapter we present the basic methodology we propose for the development of personalized predictive models applicable to the context of T1D treatment, a procedure we term *gradient method adaptation*, or GMAdapt. This procedure is highly informed by the issues and constraints encountered within the healthcare setting, particularly those which arise when attempting to apply machine learning techniques and approaches to the problems encountered in this field, and addresses them by implementing a systems engineering approach. Intelligibility and interpretability are key factors for developing system confidence and trust by stakeholders (engineers, patients, their physicians, etc.) in order for the system to be deployed effectively. Additionally, inter-subject heterogeneity can make population data derived models perform unsatisfactorily when applied to individual users, and necessitates the use of methods which enable the rapid creation of personalized models in the context of sparse amounts of subject specific data. Prolonged *run-in* periods for data collection using an unsatisfactory population model, or no model at all, are inadequate for maintaining user commitment and continuation in system use. What is needed is an application of *transfer learning* or *domain adaptation* methods which are able to improve personalized models online as new data become available. These methods should leverage population level data for model creation and initialization and be able to swiftly segue to fully individualized subject-specific models through the course of use, as well as be able to potentially track time-varying changes in the subject's underlying dynamics.

The basics of our proposed GMAdapt methodology for confronting this problem are described, evaluated, and discussed in the sections below. While we believe it is generally

applicable to many problems encountered in applications of supervised learning that face the kinds of constraints described above, the example we present is that of personalized hypoglycemia forecasting in the context of decision support systems for T1D, particularly with regard to high risk events like exercise and the overnight time period. These two tasks—creating personalized forecasting models for hypoglycemia related to exercise and overnight— provide informative test cases for how such problems will need to be approached.

In the sections below we will develop the application of GMAdapt to these tasks as described. First, the motivation for developing such methodologies is presented. Then, the overall workflow involved in the online personalization of such classifiers utilized as part of GMAdapt will be discussed, the algorithmic details of the online stochastic gradient descent based updates will follow. A series of numerical simulation based experiments, meant to establish the empirical performance of the GMAdapt updates for model personalization are then described and performed, as well as a retrospective application of the method on real-world data derived from clinical studies. The results of these simulation experiments and real-world data analysis are then presented and discussed. Finally, the chapter will conclude by comparing these results with an alternative approach utilizing hierarchical modelling approaches to GLMs which can also be used to achieve personalized models informed by population level observations, as well as discussion of the feasibility and effectiveness of using demographic and clinical data informed subgroup mappings for initialization as opposed to population models in the procedure.

3.2 Methods

3.2.1. Motivation

For tasks such as medical diagnostics, special care needs to be taken that predictive systems prove not only accurate in terms of being able to correctly categorize or estimate variables of interest,

but to be effective in contexts where decisions derived from the predictions have their own associate costs and risks. In these contexts, careful consideration of the tendencies for errors in multiple directions—over as against underestimation, false positives vs false negatives—needs to take place. One tool for direct, graphical evaluation of these tradeoffs commonly used in many clinical and engineering applications is the receiver operating characteristic curve (ROC) and the derived “concordance” or c-statistic statistic, equivalent to the area under this curve and referred to from here forward as the ROC-AUC [69].

The motivation for the methods and experiments presented below is the observation that when evaluating predictive classification models for diabetes related tasks using population level rules, the resulting ROC-AUCs obtained seemed to be attenuated by the heterogeneity present in the underlying population. To demonstrate this, we generated a Monte Carlo simulations, generating outcomes following logistic regression assumptions from subgroup aggregated populations with coefficients drawn from multivariable normal distributions with different standard deviations ranging from 0.5 to 5, and number of component subgroups (ranging from 1 to 50), performing 100 random trials for each combination. The loss in ROC-AUC which resulted from using a population level model as opposed to individualized, subgroup specific models is shown in figs. 3.1 and 3.2 below.

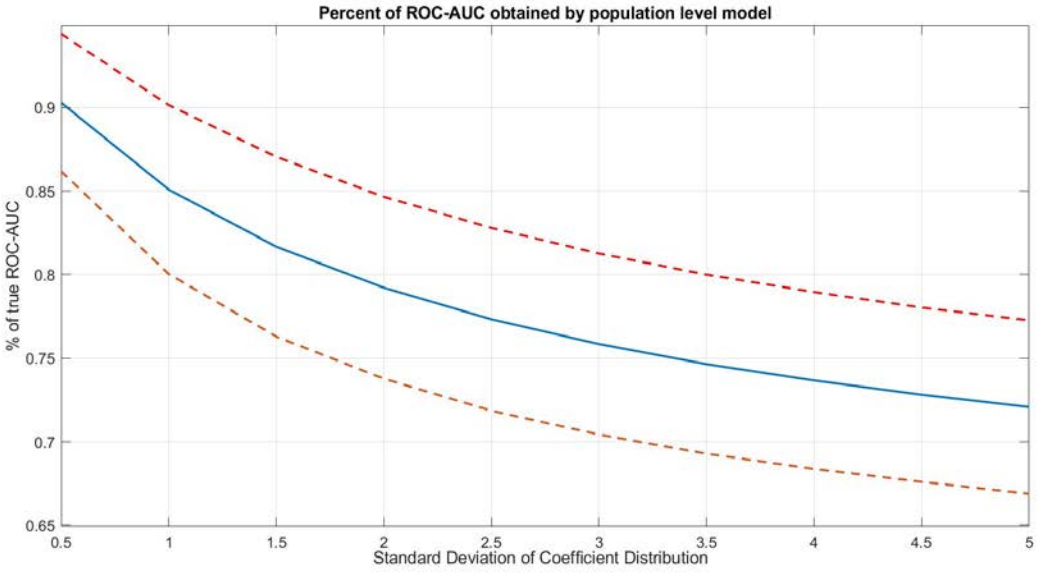


Figure 3.1: Percentage of the full subject specific model ROC-AUC obtained by using population aggregated trained classifier as a function of the variation in between subgroups.

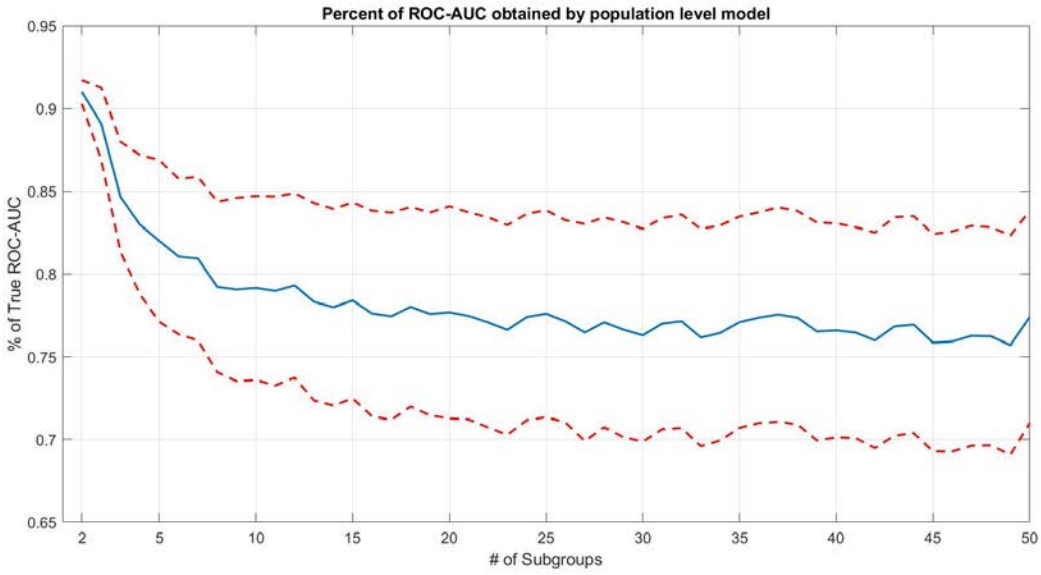


Figure 3.2: Percentage of the full subject specific model ROC-AUC obtained by using population aggregated trained classifier as a function of the number of subgroups present in the population.

As can be seen in figs. 3.1 and 3.2, there is a significant reduction in discrimination obtainable by a classifier when subgroup heterogeneity is masked by using population aggregated data for training. This phenomenon suggests that developing effective means of achieving personalized models, such as our proposed GMAdapt methodology described below, can lead to better performance for classification rules over general, population aggregated approaches when evaluated in terms of ROC derived metrics.

3.2.2 The Logistic Regression GLM population Model

In our application and analysis, we choose to implement a logistic regression classifier on our data due to its intelligibility, interpretability, and long history of use in medical applications. The model is generated using the assumptions

$$E[Y] = \boldsymbol{\pi}, \text{ b} \tag{3.1}$$

$$\log \frac{\boldsymbol{\pi}}{1-\boldsymbol{\pi}} = X\beta + \boldsymbol{\epsilon}, \tag{3.2}$$

where *Y* and *X* are the observation vector and design matrix as defined in the sections above (and specified relative to a given task), $\boldsymbol{\pi}$ is the vector of estimated probabilities, β the vector of feature variable weights, and $\boldsymbol{\epsilon}$ a vector of independent noise with distribution N(0, σI_{nxn}). Using elementary algebraic manipulations, the estimated probability that y = 1, $\hat{\pi}$, given an associated observation of the x vector of features, is given by

$$\hat{\pi} = \frac{1}{1+e^{-x\beta}} \tag{3.3}$$

Estimates of the population coefficients, $\hat{\beta}_{pop}$, are determined based on the available pooled data. Of particular relevance to our adaptation methods we note that maximum likelihood estimation of

the coefficients can be obtained by minimizing the negative cross-entropy loss function

$$L(\beta) = -\sum_{i=1}^{N} y_i \log(\hat{\pi}_i) + (1 - y_i) \log(1 - \hat{\pi}_i) \quad (3.4)$$

For the purposes evaluating GMAdapt in our data analysis and simulations below, we used the off-the-shelf Matlab® "*fitglm*" functionality to obtain logistic regression estimates of population level coefficients, β_{pop}.

3.2.3 Initialization and Personalization of Model Coefficients

The GMAdapt system is designed around the basic workflow shown in fig. 3.3. The system is initialized with coefficients determined on the aggregated, pooled population data available. At this level, feature variable determination and model selection are performed, hoping to leverage as much data as possible to determine an appropriate general model for the task. This population model is then distributed to each individual system user. As each new observation from the individual user comes in, the system advises the user's DSS of hypoglycemia risk based on the predictor variable values at the time of triggered query. At each such iteration of the system "data informed β updates" are performed.

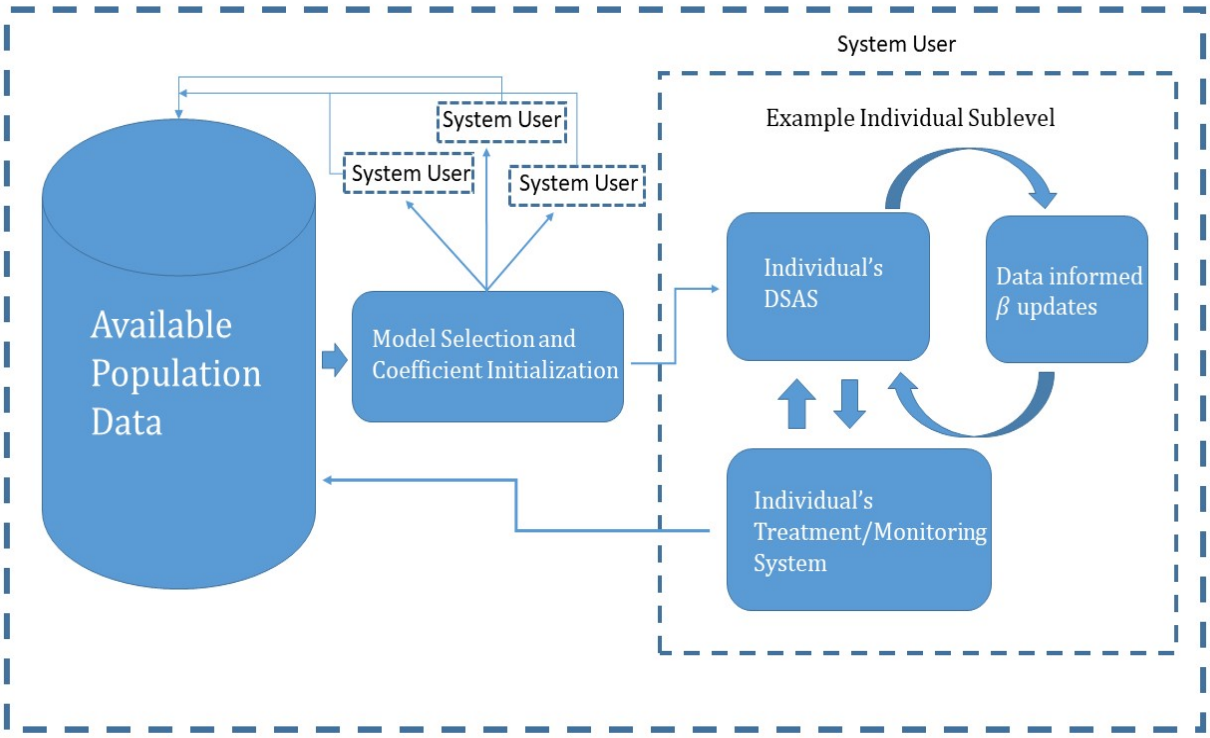


Figure 3.3: Diagram of the overall workflow for GMAdapt implementation.

These updates proceed via a single increment of stochastic (or online) gradient decent on the cross-entropy loss function [89], using the user's current β coefficients as the initialization point. This initializing and updating process is expressed in the following procedural steps:

1) Initialize system with population coefficients, $\beta \longleftarrow \beta_{pop}$.
2) On triggered query, observe associated vector of feature space variables, **x**.
3) Deliver to user's DSS the estimated probability of event, $\hat{\pi}$, based on current β.
4) Observe event window and determine the class label y. [Send total observation back to aggregate database.]
5) Update β based on set learning rate, η, and loss function gradient:

$$\beta \longleftarrow \beta - \eta \nabla L(\beta)$$

=

$$\beta \longleftarrow \beta - \eta(\hat{\pi} - y)\mathrm{x}^T$$

6) Return to Step 2.

The newly updated coefficients replace the previous coefficients for the individual subsystem and are used for prediction at the next query, the process then repeats. The observations generated can then be fed back into the database of population data in order to further refine the initial population model for new implementations/initializations as more data become available and new users are brought onto the system.

3.2.4 Numerical Simulations and Real-world Retrospective Data Analysis

We demonstrate the potential effectiveness of the GMAdapt procedure both via simulation using computational-numerical model of an underlying process that fits the assumptions and retrospectively on clinical data. Below, we describe the methodology of each of these approaches in turn.

1) Numerical Simulations

We performed numerical simulation experiments to assess the performance of GMAdapt under controlled conditions. Simulated data were generated using Matlab® functionality to approximate real application scenarios. 100 trials were performed, each with 50 virtual subjects that generated data explicitly according to the logistic regression modeling assumptions—binomial outcomes

were directly generated from sigmoid transforms of the linear predictor of the data using Matlab® functionality. For each trial a seed set of six covariate slope coefficients and one constant offset were generated from a multivariate normal distribution, and individualized true coefficients for each virtual subject were created with an additional Gaussian perturbation from this seed (zero mean, standard deviation of two). Each subject had 25 associated observations (with additive Gaussian white noise of standard deviation 0.5) represented in the aggregate pool population dataset (totaling 1250 observations). This data was used to generate model population coefficients for GMAdapt initialization. Then, a new virtual subject's data was generated using the same seed coefficients with a unique perturbation and GMAdapt (with learning rate $\eta = 0.15$) was performed on their individual data stream consisting of 100 observations with the same noise conditions as the pooled population observations. At each iteration of GMAdapt, performance of the resulting coefficients was validated on a separate dataset consisting of 1000 independent observations generated using the new virtual subject's true coefficients.

We compared performance of the predictions relative to the performance achieved by using the process's true underlying coefficients. The metrics of interest were the area under the receiver-operator characteristic curve (ROC-AUC) and detection performance with a maximum of 10% false positive rate (FPR_{10}), representing both overall performances and performance in a domain of known clinical interest.

2) Retrospective Real-world Data Analysis

In order to assess the potential effectiveness of the GMAdapt procedure in real world application, we implemented it retrospectively on data collected in clinical trials performed at the University of Virginia Center for Diabetes Technology [90]. There are two main applications we wish to

evaluate: nighttime hypoglycemia prediction and exercise related hypoglycemia prediction. In each case, we used data collected from observational studies and applied simple data cleaning and curation procedures. We applied linear interpolation to gaps in the CGM records and discarded observations with unrealistic or unusable data (coming from days with fewer than two records of carbohydrate ingestion, or fewer than two records of boluses in a day, likely indicating unreported meals or other errors, or with gaps in the signal records preventing feature or outcome variable assessment). Records of meals, insulin infusion, and CGM measurements, Fitbit® data of heartbeat, step counts, and activity level if available, as well as other clinical factors (gender, bodyweight, total daily insulin, etc.) were organized and the GMAdapt procedure was implemented for both exercise related and overnight hypoglycemia prediction as described in the corresponding subsections below.

i) Overnight Hypoglycemia Data Preparation and Analysis.

Data from the two studies were preprocessed and curated, resulting in 1106 total observations from 59 people with T1D. Subjects without any observations of nighttime hypoglycemia were excluded from analysis. The number of usable days for each subject ranged from six to 82, with a median of 17. The overall proportion of observations associated with hypoglycemic outcomes was 0.3354. The model for nighttime hypoglycemia prediction had the form of equation 3.6 below.

$$\log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot CGM_0 + \beta_2 \cdot CGM_1 + \beta_3 \cdot CGM_2 + \beta_4 \cdot IOB_6 + \beta_5 \cdot CHO_7 \tag{3.6}$$

CGM_0, CGM_1, CGM_2 are the coefficients of the zeroth, first, and second order terms in the centered polynomial interpolation of the CGM signal from the hour preceding the triggered query event (in

this nighttime hypoglycemia prediction setting, this window was 10:00pm-11:00pm). IOB_6 is the "insulin on board" at the query time as assessed by a six hour clearance curve, divided by total daily insulin (TDI), and CHO_7 is the sum of meal carbohydrates consumed in the seven hours preceding the query, divided by the individual's bodyweight in kilograms. π was the probability that a hypoglycemia would occur in the timeframe spanned by the 8 hours following the 11:00pm triggered query. In the data, labels were set as $y = 1$ if there were at least two measurements of BG<70mg/dl occurring the 11:00pm-07:00am timeframe following the query trigger point, and zero otherwise.

For each subject, the GMAdapt procedure was performed by initializing the model on the normalized population data, with the subject's own data being held out and then normalized based on the population parameters (determined excluding the subject's data), predictions and gradient updates (with learning rate $\eta = 0.15$) were then made by iterating over the subject's data. For the purpose of analysis, we looked at the ROC curves achieved by either using the specific subject-holdout population coefficients, the predictions made online through the course of adaptation, or the final adaptation coefficients retrospectively applied on each of the subjects' data streams. Particular attention was paid to the ROC-AUC and FPR_{10} metrics.

ii) Exercise related hypoglycemia data preparation and analysis

To assess GMAdapt's potential in the exercise application, data from a clinical study (GV Phase1 [90]) which had associated Fitbit® activity tracking data was used in order to approximate times of exercise and formulate a dataset suitable for testing GMAdapt in the context of exercise related hypoglycemia prediction. The trigger queries of exercise events in this analysis were determined by activity level readings greater than or equal to two as determined by the Fitbit® tracker that

continued for 20 or more minutes, with no other exercise event occurring in the previous three hours. This resulted in 873 total observations on 27 individuals (individuals with no events meeting these criteria were excluded from the analysis), with counts ranging from a minimum of three to a maximum of 71 observations (median 40) per subject. Class labels of observations were set to y = 1 if at least two readings of BG below 70mg/dl were observed in the 3 hours following the triggered query, and zero otherwise. The overall proportion of observations thus associated with hypoglycemia was 0.5178.

The basic model used for prediction of hypoglycemia associated with the exercise event had the form of equation 3.7 below.

$$\log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot CGM_{end} + \beta_2 \cdot CGM_{slope} + \beta_3 \cdot IOB_6 \tag{3.7}$$

Here CGM_{end} is the final value of the CGM readings taken before the query trigger, CGM_{slope} is the slope of the linear interpolation of the CGM signal in the hour prior to the query, and IOB_6 is the insulin on board as assessed by the six hour clearance curve, all relative to population normalized data.

The GMAdapt procedure was implemented similarly to the nighttime application above. In sequence, each individual's data were held out and population coefficients were determined on the remaining pooled data. The subject's data were then normalized according to the population parameters and the GMAdapt updating procedure was implemented iteratively (again using fixed learning rate η = 0.15) over the individual data stream. The ROC curve based analysis: comparing population, online, and retrospective predictions was then performed on the resulting data.

3.3 Results

3.3.1 Computational-numerical Simulations

Fig. 3.3 and fig. 3.4 demonstrate the performance of GMAdapt in the simulated scenarios described above.

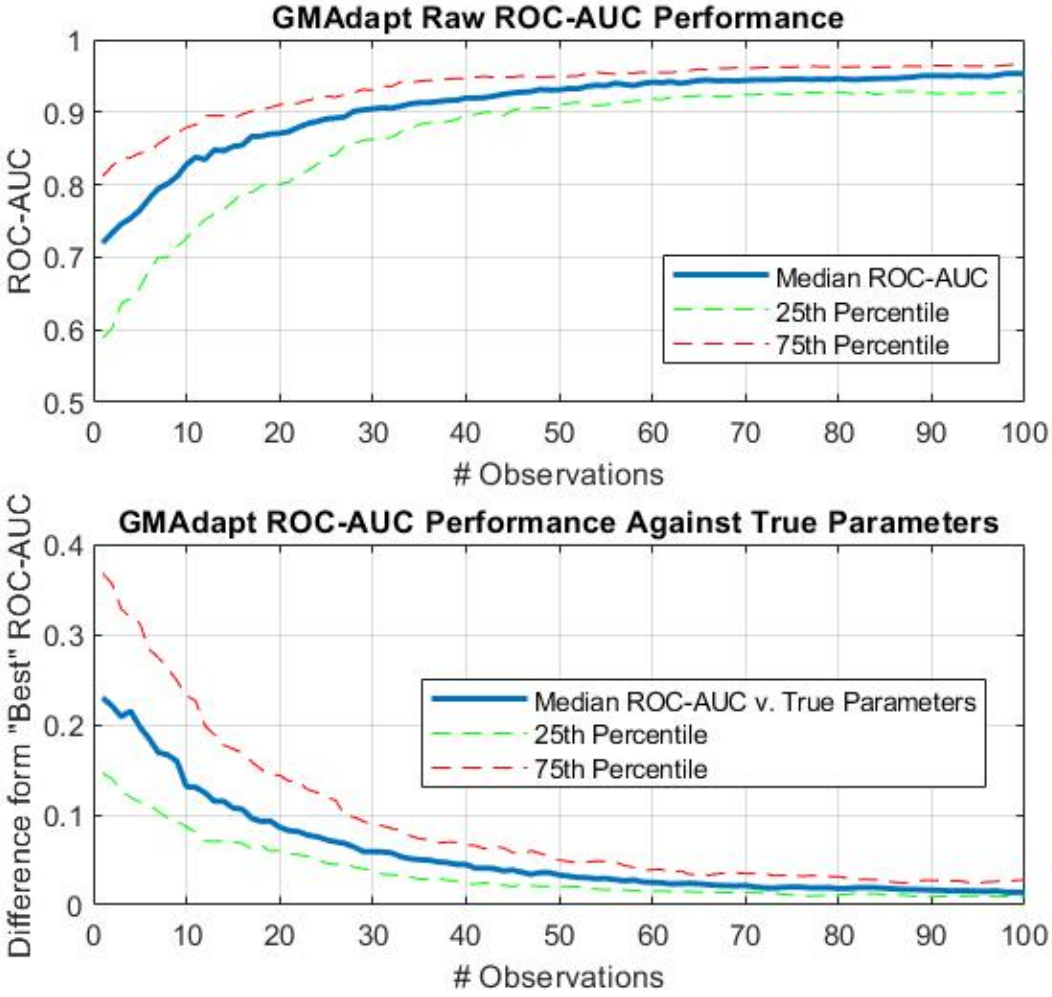


Figure 3.3 Evolution in ROC-AUC performance quartiles through the course of 100 observations, both absolute or "raw" performance (above) and performance relative to true parameter (below).

The top subplot shows the evolution of the raw performance of the coefficients obtained by the

GMAdapt algorithm on the independent validation data set over the course of adaptation. The performance began at the level of the population model for each trial (median ROC-AUC, 0.7194) and increased throughout the adaptation, achieving a median ROC-AUC of 0.9531 after 100 observations. The lower subplot shows the difference between the known true coefficients performance on the validation dataset, and those obtained over the adaptation by GMAdapt. The median difference between the true coefficient performance and the population model was 0.2297, by the end of 100 iterations of GMAdapt, it was reduced to 0.0141.

Fig. 3.4 below shows diagrams in the same format as Fig. 3.3, only focusing on the FPR_{10} . The top subplot again shows the raw performance, beginning at the population model's median of 0.3388, with final coefficients after adaptation achieving a median 0.8420 detection rate across the trials. The bottom subplot shows performance on the validation dataset relative to that obtained by the true virtual subject coefficients, again beginning at the population model performance (median 0.5512) and ending with a median FPR_{10} difference of 0.0505 from that obtained using the true virtual subject coefficients.

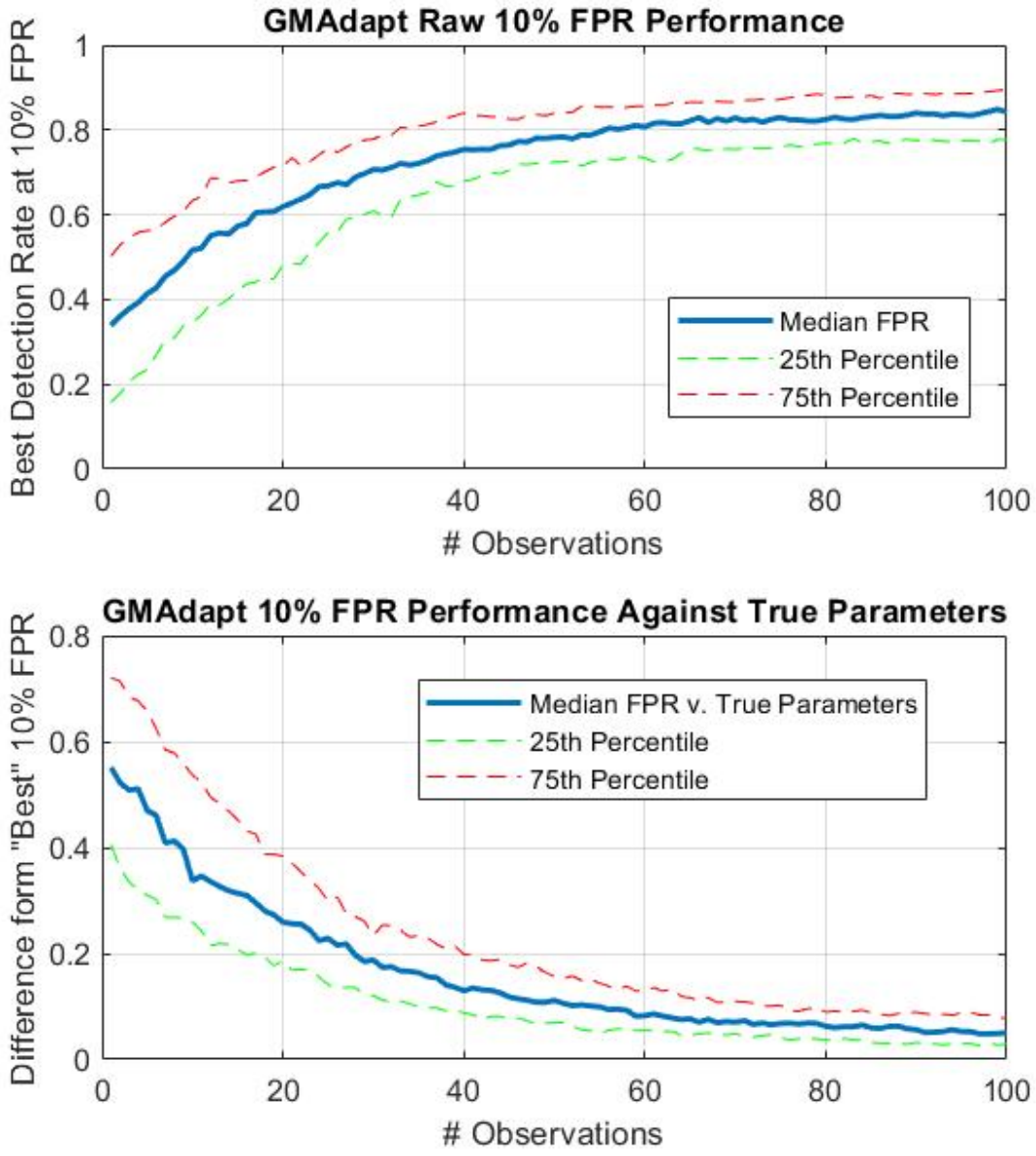


Figure 3.4 Evolution in FPR_{10} quartiles through the course of 100 observations, both absolute (above) and relative to true parameter performance (below).

3.3.2 Retrospective Real-world Data Analysis

1) Nighttime Hypoglycemia

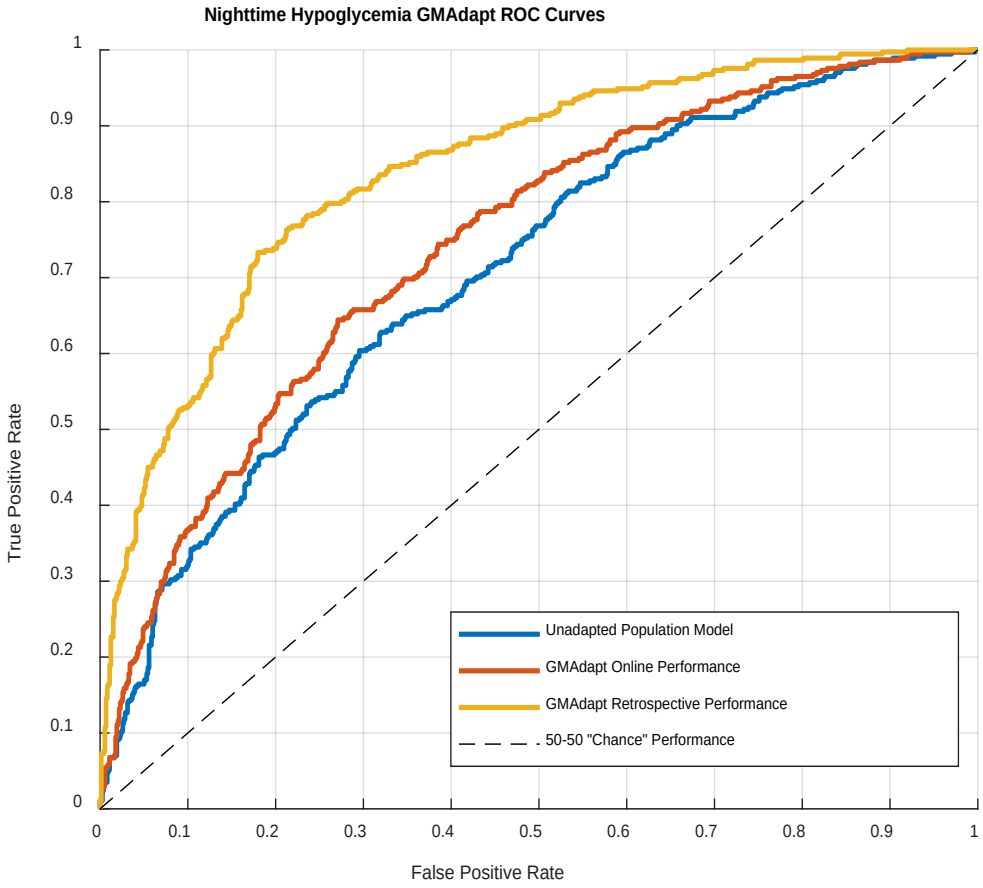


Figure 3.5: Population, online, and retrospective ROC curves for GMAdapt nighttime hypoglycemia prediction retrospective data analysis.

Fig. 3.5 presents the plots of the ROC curves obtained by the GMAdapt procedure implemented as described in the methods section above for nighttime hypoglycemia, along with comparison ROC curves. The ROC-AUC achieved by the population model, GMAdapt online through the course of adaptation, and the final coefficients obtained applied retrospectively on the data were

0.7093, 0.7439, and 0.8413, respectively, with p = 0.0028 for the comparison of online performance vs population model, and p < 2.2e-16 for retrospective performance vs the population model. FPR_{10} performances were 0.3208, 0.3666, and 0.5310, respectively, with p = 0.1453 and p = 5.401e-11 for the comparisons of online and retrospective performance vs population models respectively.

2) Exercise Related Hypoglycemia

Fig. 3.6 presents the plots of the ROC curves obtained by the GMAdapt procedure implemented on the curated exercise data as described in the methods section above, along with comparison ROC curves.

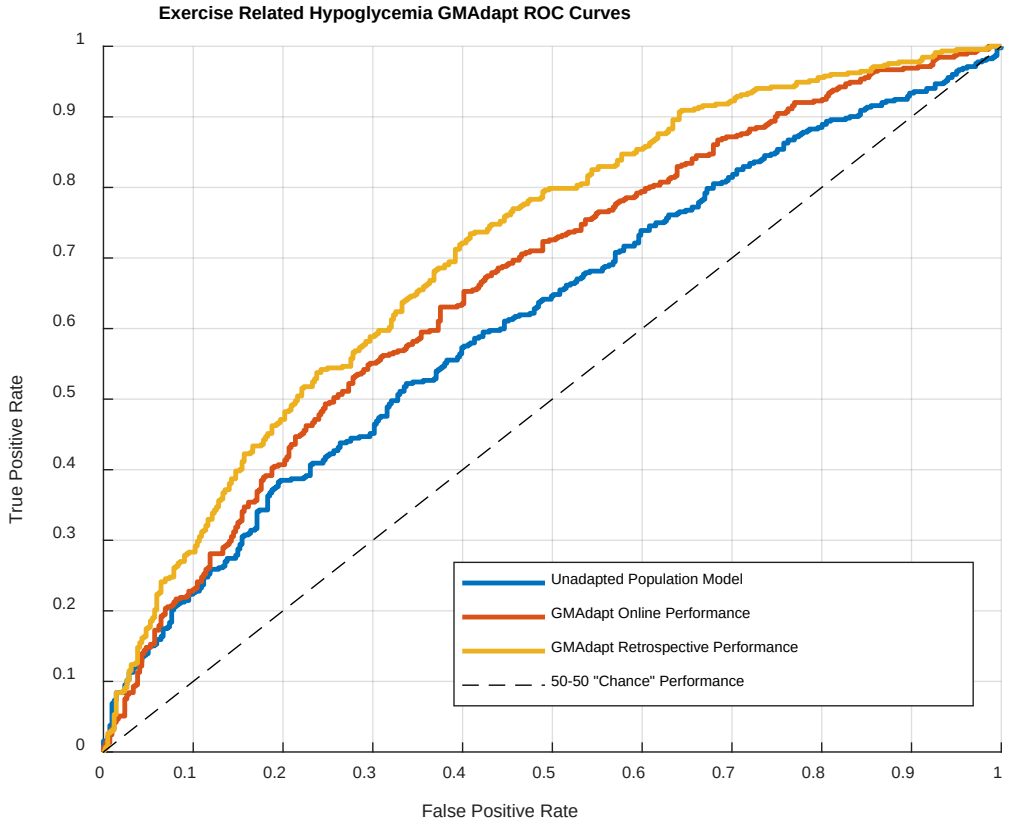


Figure 3.6: Population, online, and retrospective ROC curves for GMAdapt exercise related hypoglycemia prediction retrospective data analysis.

ROC-AUCs obtained by the population model, GMAdapt online, and GMAdapt final

coefficients applied retrospectively were 0.6165, 0.6656, and 0.7128, respectively (p = 0.03508 and p = 4.27e-06, for the comparison statistics of the online and retrospective performance vs the population model, respectively). The FPR_{10} performances were 0.2257, 0.2301, and 0.2832, respectively (p = 0.9138 and p = 0.1794, respectively, as above with the ROC-AUC).

3.4 Discussion

3.4.1 Experiment and Retrospective Data Analysis Results

Both simulation and real-world data analysis demonstrate that performance gains in terms of ROC-AUC and FPR_{10} can be obtained for logistic regression based hypoglycemia prediction systems in T1D. In the case of nighttime hypoglycemia prediction, a moderate gain in ROC-AUC was obtained during the course of the adaptation over the population model (0.0346), while a retrospective application of the final coefficients obtained achieve a more impressive gain of 0.132. Similar results for ROC-AUC were obtained in the exercise analysis (0.0491 and 0.0963 for the online and retrospective gains over population model, respectively). Limitations were such that some subjects had as few as six observations for the nighttime hypoglycemia or three observations for the exercise scenario, meaning there was little opportunity for the adapted coefficients to prove themselves for many of the subjects on the online setting, though the improvement shown be retrospective application was significant. These limitations can be overcome but future studies explicitly designed to test the GMAdapt procedure and such as currently in planning stages at the UVa Center for Diabetes Technology. Empirical data requirements—such as established "event-per-variable" (EVP) heuristics for logistic regression—for building fully individualized models from scratch can be extensive [91]. For an individual with T1D who has nighttime hypoglycemia on average once a week, the six variable classifier used above could require between 210-840 days of observation (using the 5-20 EVP heuristics) using a sub-satisfactory population model in order

to obtain enough data in order to generate a personalized model using only the individual's own data. Thus, there is clear motivation for using a process similar to GMAdapt to help a system obtain better, personalized performance from the start.

Simulation results indicate that using the GMAdapt procedure instead of the population model produces rapid gains in performance in the ROC-AUC and FPR_{10}. In as few as 20 observations, GMAdapt coefficients obtained median ROC-AUC performance lower than the true coefficients by less than 0.1, while achieving gains of 0.12 over the population model coefficients. Qualitatively similar results were obtained when focusing on performance in terms of FPR_{10}. Combined, these show superior performance of GMAdapt in the context of logistic regression based forecasting over the strategy of using an unadapted population model, or using a population model until enough data is obtained to produce a fully personalized model from scratch.

3.4.2 Comparison with Hierarchical Modeling Approaches

We consider as another possible alternative approach to GMAdapt the development of hierarchical models which included varying intercepts and varying slope structures and group level predictors unique to the individual subjects as discussed in Gelman and Hill [92]. Using the same feature variable sets above and denoting them as well as the coefficient in vectorized form by x, β, the structure of the hierarchical model takes the form given in equation 3.8 (for the i-th observation from the j-th subject)

$$\log \frac{\pi_{i,j}}{1-\pi_{i,j}} = x_{i,j}\beta + Zb_j \tag{3.8}$$

The β vector in this setting represents the overall population mean coefficients, while the b_j vector represents the subgroup (here individual) specific coefficients, with Z being a matrix encoding the appropriate varying-slope, varying-intercept structure that allows for a fully individualized models to be fitted to data.

The main question of interest here is how effectively, and at what cost, can such a hierarchical model be employed to accomplish the task of achieving a personalized logistic regression based hypoglycemia forecasting model such as GMAdapt was designed for.

To assess this, we used Matlab®'s *fitglme* functionality to developed hierarchical mixed models based on the cumulatively observable data based on the same set up and using the same data sets as in our exercise and overnight applications above. For each individual subject in the data and each of the two tasks, we initialized a logistic regression following the same structures as in the GMAdapt examples (but allowing for varying slopes and intercepts) based on overall population data (excluding the subject themselves). As new observations arrived, predictions were made and the model was update by retraining on the entire available dataset via a pseudo-likelihood method [93] (the best performing among the built in options in *fitglme*). These predictions were tracked and the resulting ROC curves were calculated as in the GMAdapt examples above. The results of these experiments and the comparison with GMAdapt are presented in table 3.1 below.

	GMAdapt Online	GMAdapt Retro	Hierarchical Model	Hierarchical Retro
Exercise ROC AUC	0.6656	0.7128	0.7183 (p = 0.000278)	0.7793 (p = 7.565e-05)

Nighttime ROC AUC	0.7439	0.8413	0.7540 (p = 0.0809)	.8316 (p = 0.2655)
Exercise FPR_{10}	0.2301	0.2832	0.2942 (p = 0.05906)	0.3739 (p = 0.06965)
Nighttime FPR_{10}	0.3666	0.5310	0.3774 (p = 0.7294)	0.5121 (p = 0.5674)

Table 3.1: Comparison of ROC-AUC and FPR_{10} performance of GMAdapt online, retrospective, and hierarchical varying-slopes, varying-intercepts model. P-values of difference between the described hierarchical models (online and retrospective performance) and equivalent respective GMAdapt performance given in parenthesis.

As can be seen, the results are mixed. For the nighttime hypoglycemia forecasting situation, the use of the online, continuously refitted hierarchical model achieves results that are not statistically different from the online performance achieved by GMAdapt. While for the exercise application, the hierarchical varying-slopes, varying-intercepts model achieves a statistically significant improvement over GMAdapt in terms of ROC-AUC, but no statistically different was obtained in terms of FPR_{10}.

These outcomes alone suggest that a personalization method based recalculating varying-slopes, varying-intercepts hierarchical logistic regression models online could potentially serve as a viable alternative to GMAdapt. However, there are some significant drawbacks to this approach. At each iteration, the entire model, across all subjects and data points, had to be recomputed. Even using the limited datasets available, this amounted to a significant increase in the computational time required—and a several orders of magnitude increase in real-time. Iterations of GMAdapt took on average 0.0004 seconds, the hierarchical modeling procedure took on average 1.02 seconds, for the exercise application. For nighttime hypoglycemia prediction iterations took 0.02

and 4.51 seconds for GMAdapt and hierarchical modeling respectively. The viability of implementing this method on mobile hardware must be questioned. Cloud based approaches may be necessary, especially as the population data grows. Increases in the available training data also pose a problem for this method. It may be necessary to use batch methods for these calculations to make it manageable, which introduces more degrees of freedom in design which would need to be assessed, whereas the GMAdapt method is *memoryless* with regard to the population data after the initial coefficients have been set.

Additionally, several other factors to do with the interaction between actions taken by the user as a result of the predictions and subsequent data, which will be addressed in detail in chapter 5 below, impede on the ability of the classifier to accurately fit data through the course of online adaptation. It is unclear how or whether the method presented their can be integrated into the hierarchical modeling approach. Nevertheless, these results indicate that the use of an update hierarchical model may prove to be a suitable method for handling problems such as those that GMAdapt addresses, provided that the time, computational, and implementation constraints, data handling, and interaction effects do not decide the issue.

3.4.2 Possible cluster based alternatives or initialization.

Examination of the core process behind GMAdapt—the leveraging of population data to develop an "average" model to use for initialization, followed by personalization—compels the question of whether or not intermediate subgroups exists between the level of the individual and the population, and how this could be used within GMAdapt or be incorporated into alternative procedures for boosting model performance in the face of heterogeneous population dynamics. In essence this breaks down into three questions which need to be answered:

1. Do clusters exist with regard different subjects' conditional distributions useful for predictions?
2. Are such clusters identifiable based on data which would be available at system initialization? Or, can we predict cluster membership using known demographic or clinical factors?
3. What do/would the existence of such clusters mean for such methods as GMAdapt?

To address these questions, we used the data from the exercise experiment described above (for which data such as age, bodyweight, HbA1c, and TDI were associated) in order to attempt to perform cluster assessments using several methods proposed in the literature [94] and implementable via the statistical and machine learning package available in Matlab®: hierarchical clustering, k-mean, and Gaussian mixture estimation. Fully individualized models were developed for each subject, and the cluster analyses was performed on the resulting coefficients.

Hierarchical clustering was implemented using Matlab® "clusterdata" functionality. All available methods within this functionality were tried to assess whether or not clustering was present. Examinations looking from 3 to 10 clusters (total number of observations 29) resulted in the maximum amount of subjects being placed into a single cluster, with singleton observations in the remaining clusters. We feel this indicates that such an approach to cluster identification is unfruitful in this application on this data. When implementing k-means and fitting Gaussian mixture models, the same dynamic repeated. This also occurred in all applications when singletons were treated as outliers and analysis was redone on the remaining subjects.

With the data available the traditional clustering algorithms applied, no feasible clusters in subject specific models were identifiable. Of course, there were limitations (only 29 subjects), and these results do not mean that further explorations of larger data sets (or other applications) will

not reveal consistent subgroup structures in individuals' models. We therefore present an outline of what an appropriate course of action would be if such subgroup structures were identified.

First, the second question posed above must be answered. That is, given that there is a subgroup structure present in the data, it must be established whether its membership in a specific subgroup an identifiable trait based on data which would be available before system initialization? This means we must establish a satisfactorily accurate method for sorting subjects into appropriate subgroups based on features such as demographic or clinical data. If that is impossible it is difficult to see how an effective method of either integrating subgroup information into GMAdapt or developing alternative methods based on this information could be developed.

Assuming that this first hurdle is overcome, the next question is what to do with this information. If subgroup level models can be developed which account for most of the variability and heterogeneity that is observed, and there is no gain of using fully subject personalized models over these subgroup level models, then the applications of methods like GMAdapt would be limited to things like tracking the time-varying phenomena that may persist. However, if there are still meaningful gains to be had by using a fully personalized model over the subgroup level model, it would be a simple matter of identifying the individual's subgroup average parameters and initializing on those as opposed the population level average, but otherwise implementing GMAdapt in an identical manner.

3.5 Conclusion

The above methods and experimental results present a means, GMAdapt, of achieving personalized predictive models for hypoglycemia forecasting in T1D. While the performance both in numerical simulations and applied retrospectively to real-world clinical data are encouraging, there are still questions which need to be answered in order for GMAdapt to achieve effective real world implementation. In the following chapter 4 we present a modification of a simulation methodology, the “net effect” method, and in chapter 5 we will present proposal for integrating this newly updated net effect methodology together with GMAdapt to overcome some of the confounding issues which may hamper online adaptation methods in practice if left unaddressed.

Chapter 4

4.1 Introduction

In this chapter we present a simulation methodology for *retrodicting* signals based on a dynamic model of the underlying generating processes and known system inputs. This work is an extension of simulation methods developed by Patek et. al [4] based on real-world observations of data intended to capture the variability which is seen in real world CGM traces but lacking in results generated by simulation platforms such as the UVa/Padova simulator. The extension uses system identification methods to fit some critical parameters of the underlying model in order to overcome issues which limit the effectiveness of the procedure in replaying hypothetical scenarios when assessed against simulation data. This extensions of the net effect methodology will be applied in chapter 5 in order to address the impact of the introduction of confounding medical interventions on the GMAdapt procedure for model personalization when used for hypoglycemia prediction in the context of a decision support system. Before explaining and evaluating the methodology in detail, a brief background of the basic method and the motivations for the proposed updates is presented in the subsections immediately below.

4.1.2 Net Effect Background

Patek et. al [4] proposed using a deconvolution based methodology together with a linear time-invariant (LTI) model of BG and insulin dynamics to extract a forcing "net effect" signal representative of the impact of unmodeled phenomena in order to achieve simulations which reproduced the variability seen in real-world CGM traces but that was lacking in platforms such as the UVa/Padova simulator. This extracted net effect signal could be fed back into the LTI model together with the known insulin record in order to reproduce the observed BG trace. This simulated

BG trace provides insight into the possible effect of alterations in insulin treatment for the specific patient in the timeframe of the observed data.

In practice, this net effect simulation methodology has seen use by Kovatchev et al. [95] to assess the tolerable bounds for sensor errors to allow for the non-adjunctive use of CGMs in place of self-monitored blood glucose (SMBG) when making insulin treatment decisions. However, there are some limitations and issues with this methodology. First, as currently described and implemented, the net effect signal is estimated and enters into the LTI model as an additive input through the initial meal compartment in the underlying model of glucose-insulin dynamics. But the net effect signal is meant in principle to account for *all* potential extra model sources of BG variation, and hence it makes more sense intuitively and conceptually for it to enter directly as an additive disturbance in the blood glucose compartment. While this approach is still incomplete (BG variation may be due to non-additive phenomenon, e.g. increase in insulin sensitivity), it is more general than the original meal pathway [4]. Additionally, since the original net effect is calculated from the insulin record and BG measurements only, simulations or "replays" are necessarily limited to alterations in insulin therapies, while adjustments in meals or hypotreatments (meals without associated insulin boluses—used to prevent or treat hypoglycemic events) cannot currently be implemented in the simulation procedure without altering the net effect signal itself, and hence cannot be evaluated directly using the methodology in its current form.

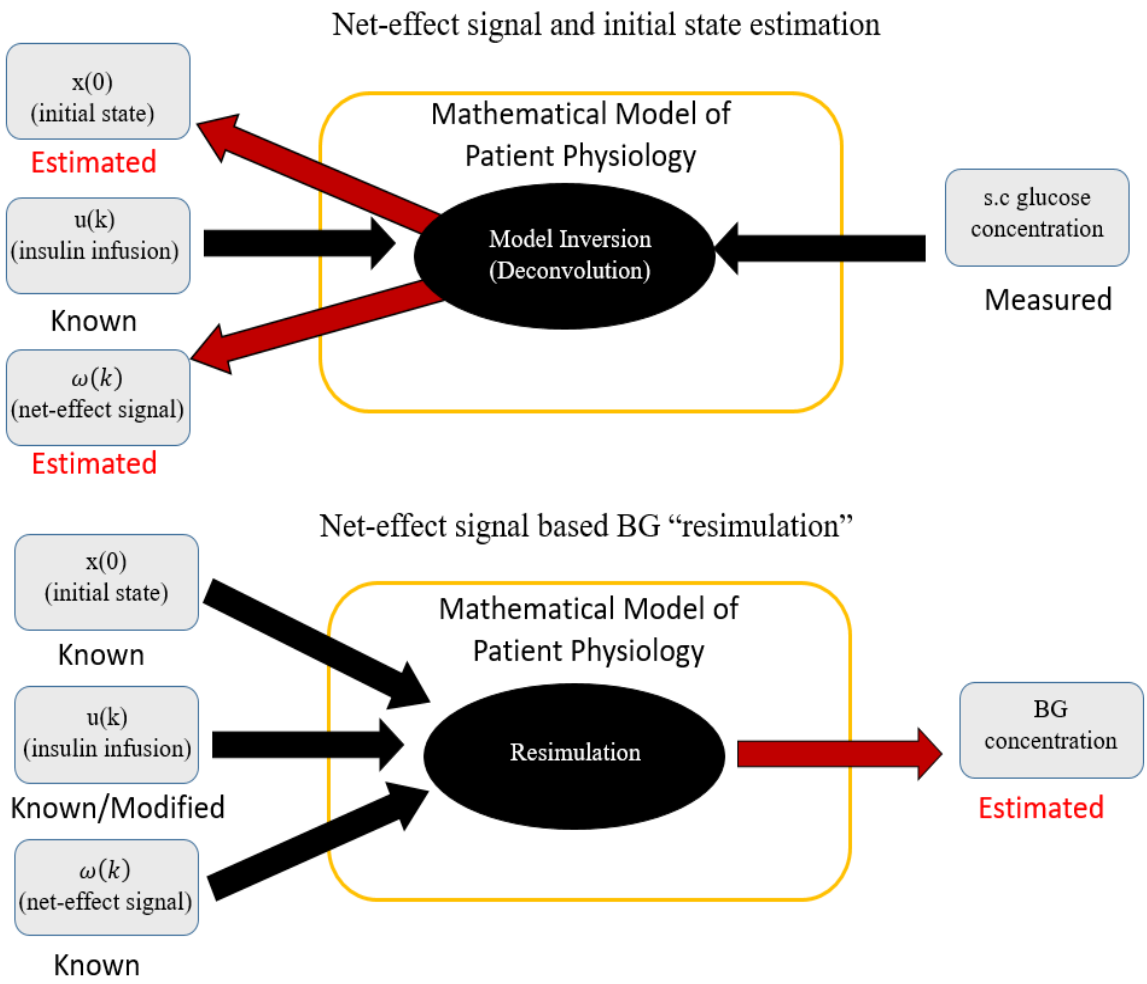


Figure 4.1: A schematic diagram of the net effect simulation procedure. In the top diagram, a visual representation of the net effect signal calculation as described in [4] *is shown. Below, the resimulation procedure, in which the net effect signal is used to "replay" the data to estimate BG traces with modified insulin is demonstrated.*

More broadly, and as pointed out by Vettoretti et. al. [96], the construction of such hypothetical "replays" leads directly to the question of the "domain of validity" of this simulation procedure. As discussed in that manuscript, the notion of "replaying" real-world data is inherently problematic. Because the net effect is meant to represent the variability in BG derived from sources

that are not directly incorporated into the underlying LTI model, it is assumed to be independent of changes in the modeled phenomena (e.g. changes in insulin dosage) and to be fixed over the timeframe on which the deconvolution is calculated. There are many good reasons to call these assumptions into question. The original net effect signal itself was calculated based on knowledge of insulin records, and thus the assumption that it is independent of changes in the insulin record is questionable. Assessing the extent to which this lack of independence between the net effect calculation and the system inputs influences the BG signal reconstruction under differing insulin treatments would require a comparison of the net effect based simulation results with actual observed data from the various scenarios under consideration. However, since the net effect signal is specific to a given timeframe, such data is impossible to obtain in reality—we cannot "replay" a real-life day of data collection under an alternative scenario in order to validate the simulation. As a substitute, Vettoretti et al. used the UVa/Padova simulator in order to produce data which would allow the net effect replays under altered therapies to be compared against nominal "ground truth" obtained by rerunning the alternative scenarios in simulation. Such an approach unfortunately is limited by the UVA/Padova simulation platform's imperfect representation of the real-world phenomena under study, the very reason replay techniques such as the net effect simulation procedure were explored.

For net effect simulation replays to be used as an effective method for the development of T1D treatments and technologies, these problems need to be addressed, preferably with meal inputs explicitly incorporated into the deconvolution procedure (fig. 4.2). The purpose of this chapter is to propose an updated net effect methodology developed in part to account for some of these issues. This updated net effect shifts the deconvolution based calculation of the forcing signal from the meal compartment state directly to the blood glucose compartment state and incorporates

known meal and hypotreatment data in the deconvolution calculation. Additionally, it introduces both triangular meal and insulin models into the underlying LTI system—the subcutaneous oral glucose minimal model (SOGMM)— in place of the current sequential insulin and meal compartment based SOGMM, in order to more accurately capture real-world glucose-insulin dynamics. Finally, an initial step is added to the algorithm computing the net effect signal itself. This initialization step is a system identification of some relevant parameters on the target data in order to individualize or tune the SOGMM model before the net effect is calculated, limiting the impact of model mismatch between the appropriate individual and the population level SOGMMs on the net effect signal estimation.

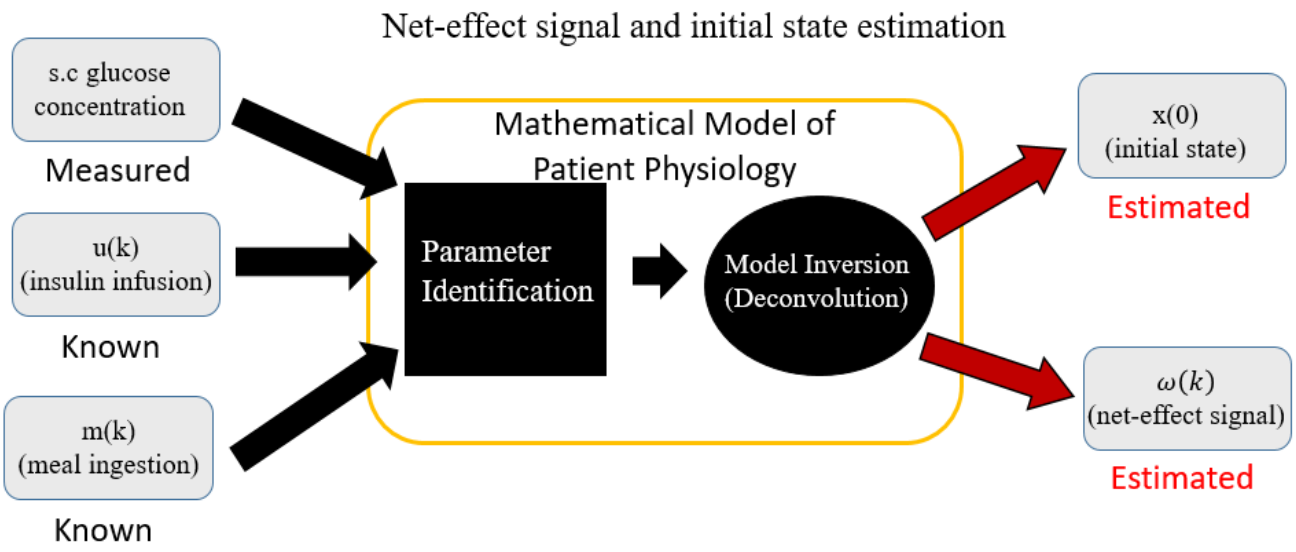


Figure 4.2: Derivation of the net effect signal including meal records and our proposed initial system identification.

The proposed methodology for implementing these changes will be presented in the following section, and we hope that the motivations for the former two changes (i.e. moving the point of input of net effect signal in the LTI from the meal compartment to the blood glucose compartment, and using triangular insulin and meal sub-models) are straight forward enough that

their brief statements below are sufficient. However, we would like to expound a bit more on the motivation for the last change—the addition of a model parameter identification step prior to the net effect calculation. The net effect is meant to capture the extra-model variability in BG signal. Under the assumption that the model in fact accurately represents BG-insulin dynamics (and, additionally, meal ingestion dynamics in our current implementation), this variability can have many sources: measurement noise, misreported data, unmodeled phenomena such as changes in insulin sensitivity, etc. However, there is also the possibility that the discrepancy between model-predicted BG and the actual observation is due to the model used itself. For example, the linearized, discretized sequential SOGMM used by Patek et.al was implemented with population level parameters for many components, while the UVa/Padova simulator consists of 100 unique parameter sets representing the individual virtual adult subjects. Thus, even if operating under the idealized assumption that there is no discrepancy between the model and the observed data due to the factors the net effect is meant to account for, the signal will attempt to compensate for the different dynamics due to these differing model parameters for each of the virtual subjects. In fig. 4.3 below, an example case displaying the tendency of the net effect signal to compensate for model discrepancies, as opposed to unmodeled phenomena, based on simulator data is presented.

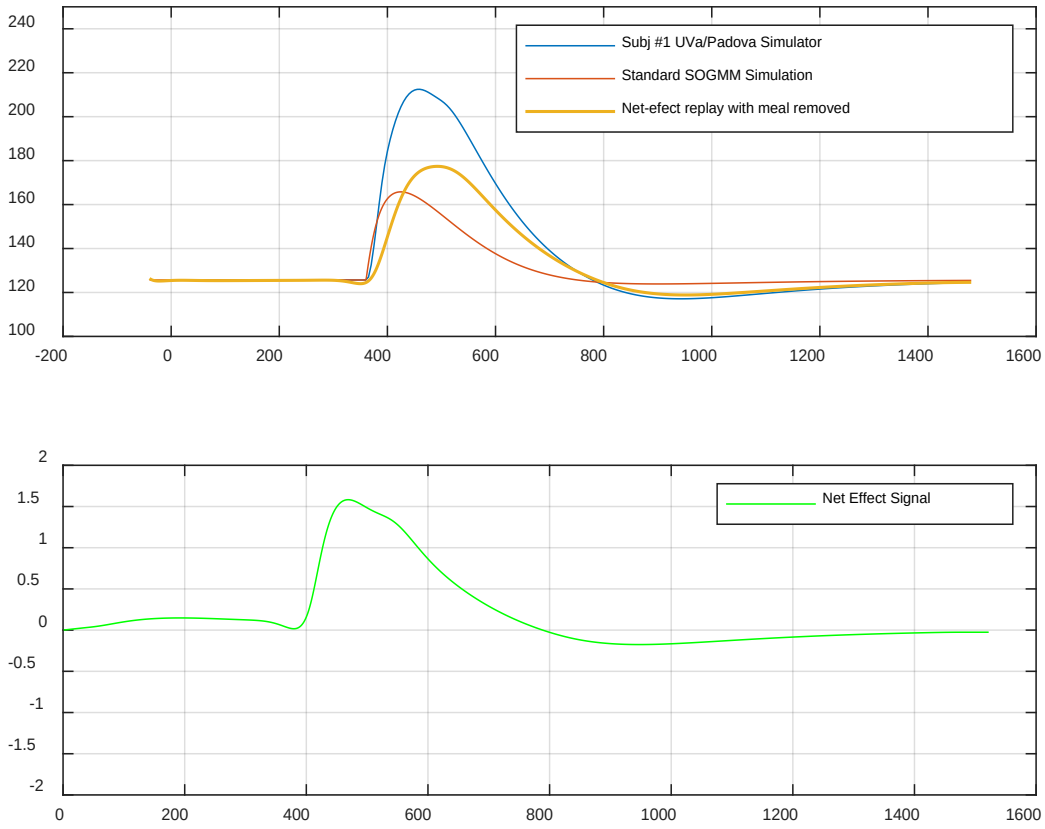


Figure 4.3: An example of the net effect signal based on a single 40g meal and bolus/basal insulin, the "open loop" results from both the SOGMM and Uva/Padova simulator for virtual subj #1, and the "replay" of the SOGMM model with the calculated net effect "fed" back into the model, but with the meal and bolus removed. The net effect signal here merely reflects model discrepancies and not any unmodeled phenomena, of which there are none in this simulated environment. The final result is an unrealistic simulation of the system with no disturbance (yellow), which should result in steady-state behavior.

Differences in the way in which the underlying model and the true plant (here the UVa/Padova simulator) handle the meal lead to a net effect signal that compensates for the gap

between the BG traces, which is here due to the manner in which the model handles the meal and insulin dynamics, and not some exogenous unmodeled factor as intended. This results in an unrealistic “replay” when the meal and insulin bolus are removed, essentially reproducing a “phantom” meal because of the mismatched meal time constants and glucose gain factors. When some parameters related to meal and insulin dynamics are identified before calculation, this phantom effect is attenuated, leading to a more reasonable approximation of the expected steady-state behavior which should be observed when no insulin boluses or meals are given, as demonstrated below in fig 4.4.

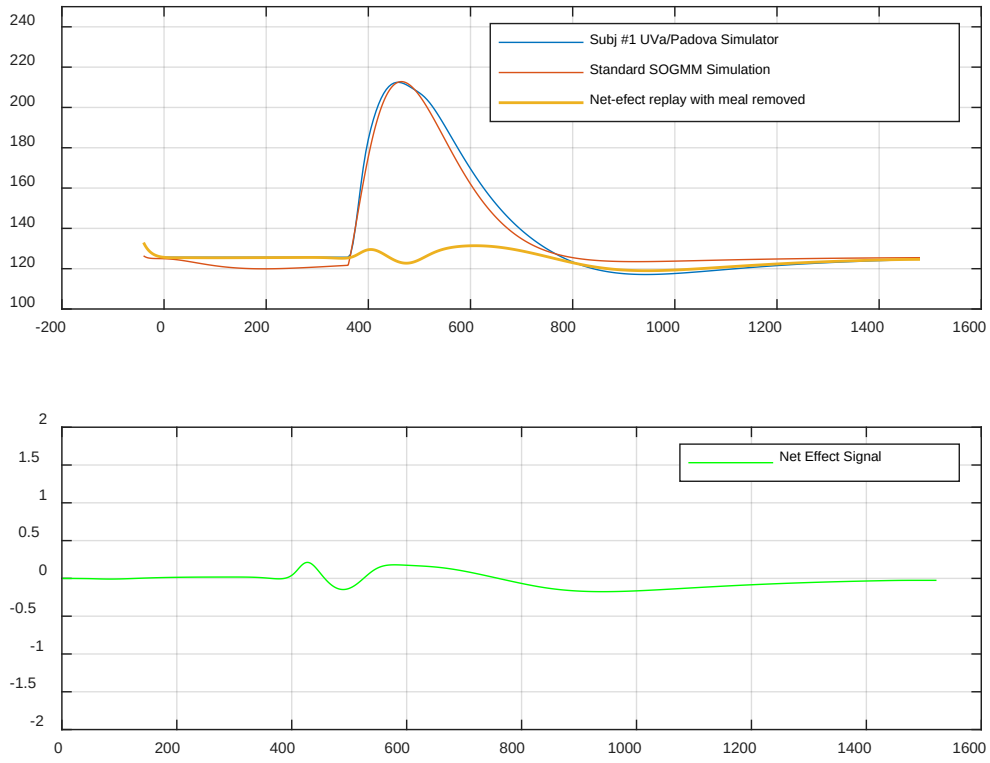


Figure 4.4: The same scenario as Figure 4.3, but using a pre-identified SOGMM model to account for model discrepancies. This results in a much more realistic resimulation of the system with the meal and bolus removed—mild oscillations around the steady-state, deviating less than 10mg/dl

from fasting BG.

In the following sections we will present our methodology for updating the net effect simulation procedure. Additionally, we will perform experiments and analysis using simulation conducted with the UVa/Padova simulator to assess the possible domain of validity for this new, updated version. Finally, we will discuss the results of these experiments and present a potential clinical application on real-world data as an example for its use in the development T1D treatments going forward.

4.2 Methods

4.2.1 The Updated Subcutaneous Oral Glucose Minimal Model "SOGMM"

The original net effect procedure was centered around a linearized, discretized version of the SOGMM model of glucose-insulin kinetics, derived from the standard minimal model as presented by Dalla Man et al [41]. These nonlinear dynamics can be represented by the equations

$$\dot{G}(t) = -\Big(S_g + X(t)\Big) \cdot G(t) + S_g \cdot G_b + (R_a(t))/V_g \tag{4.1}$$

$$X(t) = -p_2 \cdot X(t) + p_2 \cdot S_I(I(t) - I_b) \tag{4.2}$$

where $G(t)$ is the plasma glucose concentration (mg/dl), $I(t)$ the plasma insulin concentration (mU/L), $R_a(t)$ the plasma glucose rate of appearance (mg/min), and $X(t)$ the proportion of insulin in the remote compartment (1/min). On top of this "core" model of plasma glucose-insulin

dynamic, the SOGMM includes subcutaneous insulin and gastrointestinal carbohydrate transport sub-models, using sequential, compartmental models.

In our updated implementation, these two sequential sub-models—the gastrointestinal oral carbohydrate model and the subcutaneous insulin kinetics sub-model— are altered to triangular models, allowing for direct diffusion paths from their corresponding first compartments to the BG or plasma insulin compartments, respectively. In order to change sequential meal model, where meal inputs pass from the first comportment Q_1 through to the second Q_2 and then appear in the bloodstream, into a triangular model, all that is required is for there to be a properly scaled term directly linking the first compartment to the blood glucose state. This triangular model is written as

$$\dot{Q}_1(t) = -(k_{q_1} + k_{q_{12}})Q_1(t) + u_m(t) \tag{4.3}$$

$$\dot{Q}_2(t) = k_{q_{12}}Q_1(t) - k_{q_2}Q_2(t) \tag{4.4}$$

$$R_a(t) = \frac{k_{q_1} \cdot f}{V_G}Q_1(t) + \frac{k_{q_2} \cdot f}{V_G}Q_2(t) + \frac{\omega(t)}{V_G} \tag{4.5}$$

where k_{q_1} is a rate of appearance term associated with the first compartment (new to the sub-model), $k_{q_{12}}$ is the transport term between compartments (analogous to k_τ in the original, sequential model) and k_{q_2} is the rate of appearance term of the second compartment (analogous to the k_{abs} parameter in the original SOGMM). A similar triangular model can be implemented with insulin kinetics, leading to a triangular insulin sub-model of the form:

$$\dot{I}_{sc1}(t) = -(k_{c_1} + k_{c_{12}})I_{sc1}(t) + u_i(t) \tag{4.6}$$

$$\dot{I}_{sc2}(t) = k_{c_{12}}I_{sc1}(t) - k_{c_2}I_{sc2}(t) \tag{4.7}$$

$$\dot{I}_p(t) = -k_{cl} \cdot I_p(t) + k_{c_1}I_{sc1}(t) + k_{c_2}I_{sc2}(t) \tag{4.8}$$

These triangular sub-models allow us to more accurately represent the kinetics of meal and insulin in the body, conceptually by permitting both "fast" and "slow" carbohydrate and insulin action. This is especially useful for our system identification procedure, since the added flexibility of fitting both fast and slow meal parameters allows us to more readily match the dynamics of real-world BG signals due to varying meal composition or other factors.

For practical purposes, implementation of the triangular model for net effect calculation uses the matrix

$$\tilde{A} = \begin{bmatrix} -S_G & -G_b & 0 & 0 & 0 & 0 & \frac{k_{q_1} \cdot f}{BW \cdot V_G} & \frac{k_{q_2} \cdot f}{BW \cdot V_G} \\ 0 & -p_2 & 0 & 0 & \frac{p_2 \cdot S_I}{V_I \cdot BW} & 0 & 0 & 0 \\ 0 & 0 & -(k_{c_1} + k_{c_{12}}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{c_{12}} & -k_{c_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{c_1} & k_{c_2} & -k_{cl} & 0 & 0 & 0 \\ k_{sc} & 0 & 0 & 0 & 0 & -k_{sc} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_{q_{12}} - k_{q_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{q_{12}} & -k_{q_2} \end{bmatrix} \tag{4.9}$$

while following the standard discretization and net effect deconvolution calculation procedure used in the previous implementation [4]. In addition to these modifications, further input terms to account for generally known inputs beyond insulin records, such as meals or hypoglycemia treatments, are added. Furthermore, the net effect signal now enters the system directly in the differential BG concentration state, as opposed to passing through the meal subsystem and entering in the first gut compartment as before. The resulting (*linearized*) system will have the form

$$\dot{x}(t) = Ax(t) + B_i u_i(t) + B_m u_m(t) + G\omega(t) \tag{4.10}$$

$$y(t) = Cx(t) \tag{4.11}$$

Where, u_i and u_m represent the differential insulin and meal records, respectively, while $B_i = [0\ 0\ 1\ 0\ 0\ 0\ 0\ 0]'$, $B_m = [0\ 0\ 0\ 0\ 0\ 0\ 1\ 0]'$, and the net effect input matrix is G $=[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]'$. With a state space the same as used in [4].

4.2.2 Using model identification for parameter personalization in the net effect simulation

In its original formulation, the net effect procedure used population level parameters , except for bodyweight (BW), insulin sensitivity, S_I, and fasting blood glucose, G_b, each of which were given subject specific values. These were either directly assessed from medical records (BW) or calculated via heuristic methods based on total and basal insulin delivery for S_I, or HbA1c, for G_b. To update these methods we propose to use a nonlinear least-square fitting methodology to identify S_I, f, k_{q_1} , $k_{q_{12}}$, and k_{q_2} on the data directly. These parameters were chosen so as to allow close model fit to meal dynamics, therefore limiting impact of meals on the determination of the

net effect signal. A priori local identifiability of these parameters was confirmed via computer algebra methods (DAISY [97]). Before the deconvolution calculation of the net effect signal, we identify these parameters on the BG, insulin, and meal data itself via a least-square best fit of the BG trace against the "open loop" signal generated the triangular SOGMM model using Matlab®'s *lsqnonlin* functionality implementing the "trust-region-reflective" algorithm on default settings. S_I and f are identified for each day of data, while the triangular meal parameters—k_{q_1} , $k_{q_{12}}$, and k_{q_2} — can be fitted for individual meals by augmenting the system with additional sub-compartments for each additional meal in a straightforward extension of the system matrices and state space vector.

4.2.3 Simulation experimental set-up

In order to evaluate our updated net effect simulation methodology, we propose several experiments to be performed using data generated using the UVa/Padova type-1 diabetes simulator. The use of the simulator allows us to retrieve counter-factual ground truth of the different experimental scenarios for comparison with the data generated from the net effect simulations, something impossible to accomplish using real-world data. Across each of the 100 virtual adult T1D patients present in the simulator the following experiments were performed:

1.) Simulation of altered meal sizes

Each of the virtual subjects received one meal consisting of 0.7gCHO/kg of bodyweight consisting of fast-oral carbohydrates at noon under their standard insulin therapy, correcting to range and with fixed insulin sensitivity across the day. The net effect was calculated on this data and used to replay increments of 20% reductions and increases in meal sizes. The resulting BG traces

generated by the net effect resimulation are compared with post-prandial (12hr) root mean-square error (RMSE) to assess performances across the population.

The post prandial RMSE is useful to give insight into the average error between the BG generated by the net effect simulation and the "true" BG trace obtained from actually rerunning the scenario through the original plant—the UVa/Padova simulator—rerun with the same adjustments. We focus on the *post-prandial* RMSE specifically, since both the UVa/Padova simulator and the net effect resimulation will be in steady-state until the meal/insulin disturbance arrives at noon, resulting in significant underestimation of RMSE if this period is included in the analysis. In addition, evaluations of the performance of the net effect simulation against the true dynamics were assessed via percentage time in regions defined by the *Clark error grid* [98], a metric used in T1D research to assess the trade-off between accuracy and risks inherent to different BG ranges. The Clark error grid is an often used measure in diabetes care which partitions the planar segment created by comparison of BG estimates with reference values (in our case, compares net effect simulations of scenarios to the ground truth obtained from the UVa/Padova simulator) into 5 zones representative of clinical performance: A, B, C, D, E. It is designed to go beyond numerical accuracy, given that only deviations that may lead to different clinical decisions are relevant to assessing performance with regard to treatment of diabetes (e.g. a prediction of 130mg/dl for a true value of 110mg/dl is inconsequential, while the same 20mg/dl deviation at an 80mg/dl reference level can have significant consequences for treatment decisions). Ideal performance would result in a high percentage of estimates falling in zone A or zone B, with minimal time spent in the other zones.

2.) Simulation of removed meals of varying sizes

Similar to experiment 1, each of the 100 adult virtual subjects received a meal at noon of the simulation day under their standard insulin bolus therapy with fixed insulin sensitivity (no dawn effect or intra-day variation). The meals corresponded to the adjustments made in the previous experiment, ranging from 0.14gCHO/kg-BW to 1.4gCHO/kg-BW of fast-oral carbohydrates. Then, the meal and accompanying bolus were removed entirely using the net effect resimulation. The ability of the BG trace simulated by the net effect to match the comparator data from the UVa/Padova simulator with no meal was again assessed across the population using post-prandial RMSE and Clark error grid analysis (CEGA).

3.) Simulation of bolus adjustments for 0.7g/kg noon meal

The same initial set-up as experiments 1 and 2 is used. Reductions and increases in the meal bolus, ranging from a 30% decrease through a 30% increase in 10% increments of the original bolus are simulated using the net effect signal and compared to the ground truth obtained via rerunning the altered scenario the UVa/Padova simulator. Again, the base metrics of comparison were the post-prandial RMSE and CEGA.

4.) Simulation of basal adjustment

To assess the ability of the proposed net effect methodology to accurately reproduce the impacts of basal insulin adjustments, alterations of basal insulin ranging in 10% increments from a 50% reduction to a 50% increase were resimulated via the net effect and compared against the results obtained by rerunning the same scenarios in the UVa/Padova simulator. To assess basal performance, a full simulated day of data was generated, with three meals of size 0.4g/kg-BW, 0.8g/kg-BW, and 0.7g/kg-BW given at 07:00, 12:00, and 18:00 hours, respectively. RMSE and

CEGA performance were calculated over the *entire* day, as opposed to the merely the post-prandial period, since this is both the standard for basal assessment and the steady-state issues which produced over-optimistic results for the other experiments were not a significant source of overestimation of performance in the basal adjustment case.

4.3 Results

The results in terms of RMSE performance across the four experimental scenarios are presented in the graphs (Fig. 4.3-4.6) below.

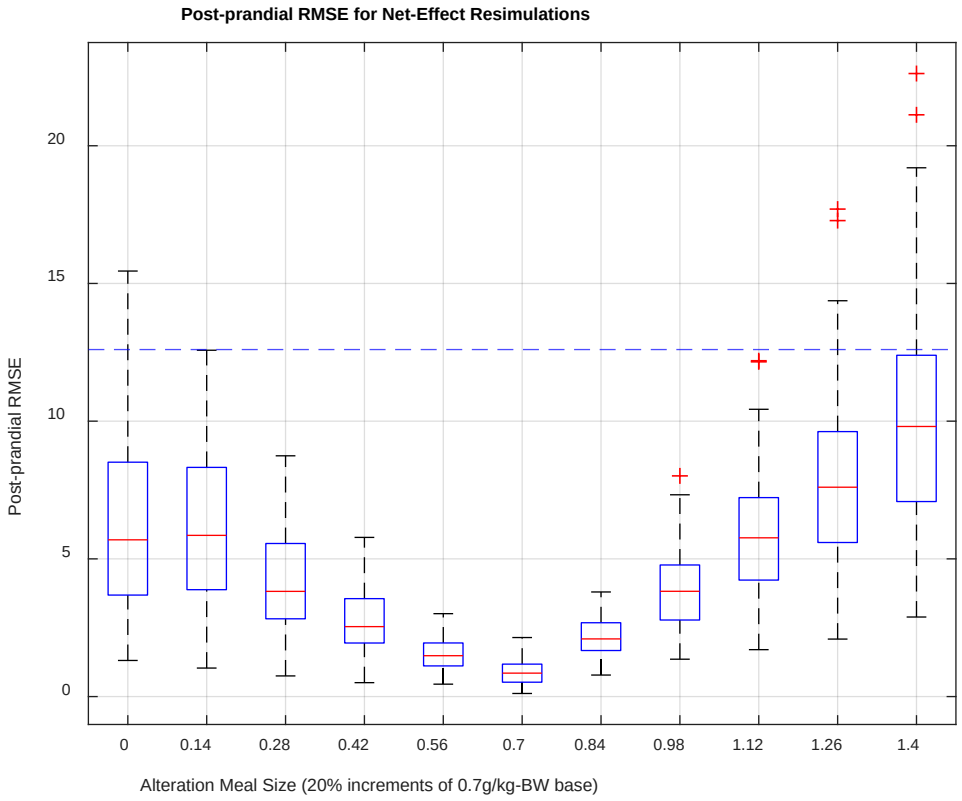


Figure 4.3: Post prandial RMSE results for net effect resimulations of altered meal sizes for singal meal.

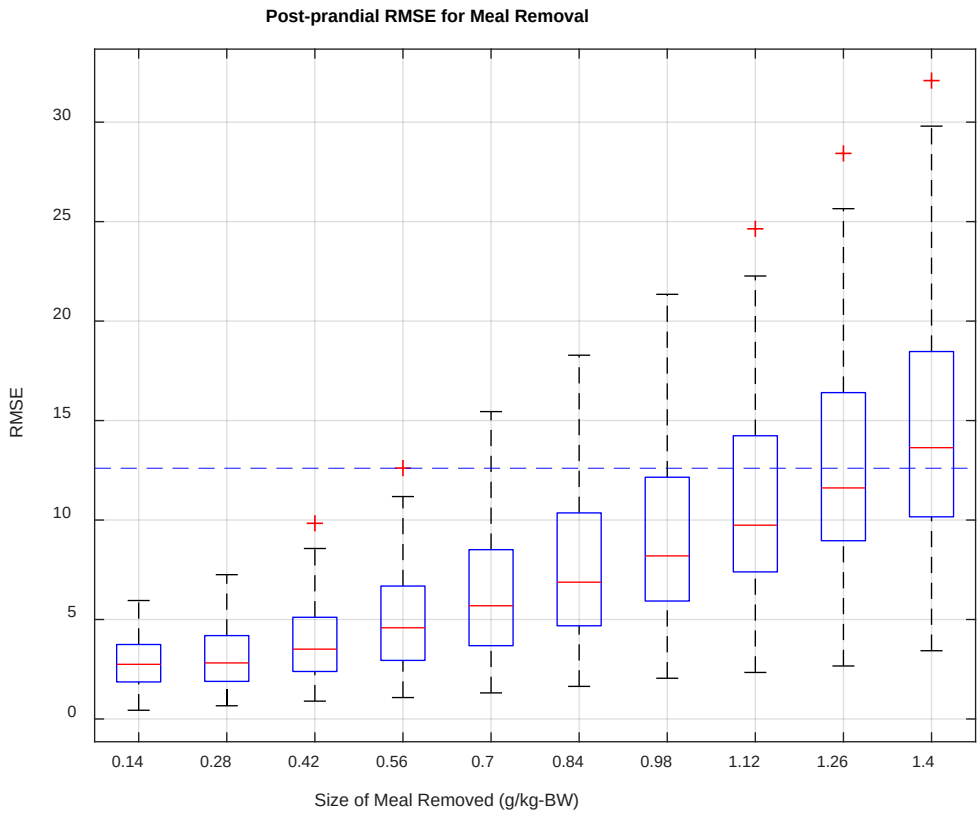


Figure 4.4: Post prandial RMSE results for net effect resimulations of total meal removal for singal meal of varying sizes.

In terms of CEGA, for simulation experiment 1 all meal size changes below an 160% increase resulted in 100% zone A or B performance. For increases 160%-200%, zone D performance was observed, but never in excess of 2.3%. There was no zone C or E performance was observed.

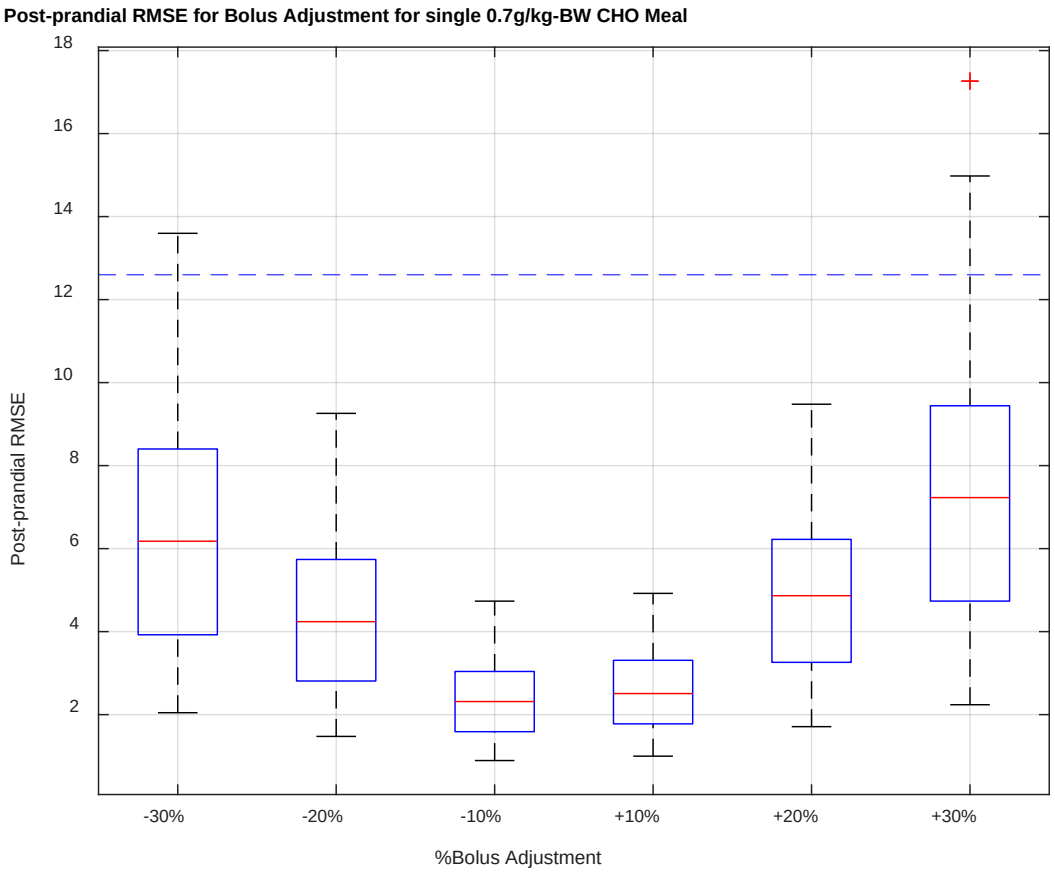


Figure 4.5: Post prandial RMSE results for net effect resimulations of altered bolus sizes for singal meal.

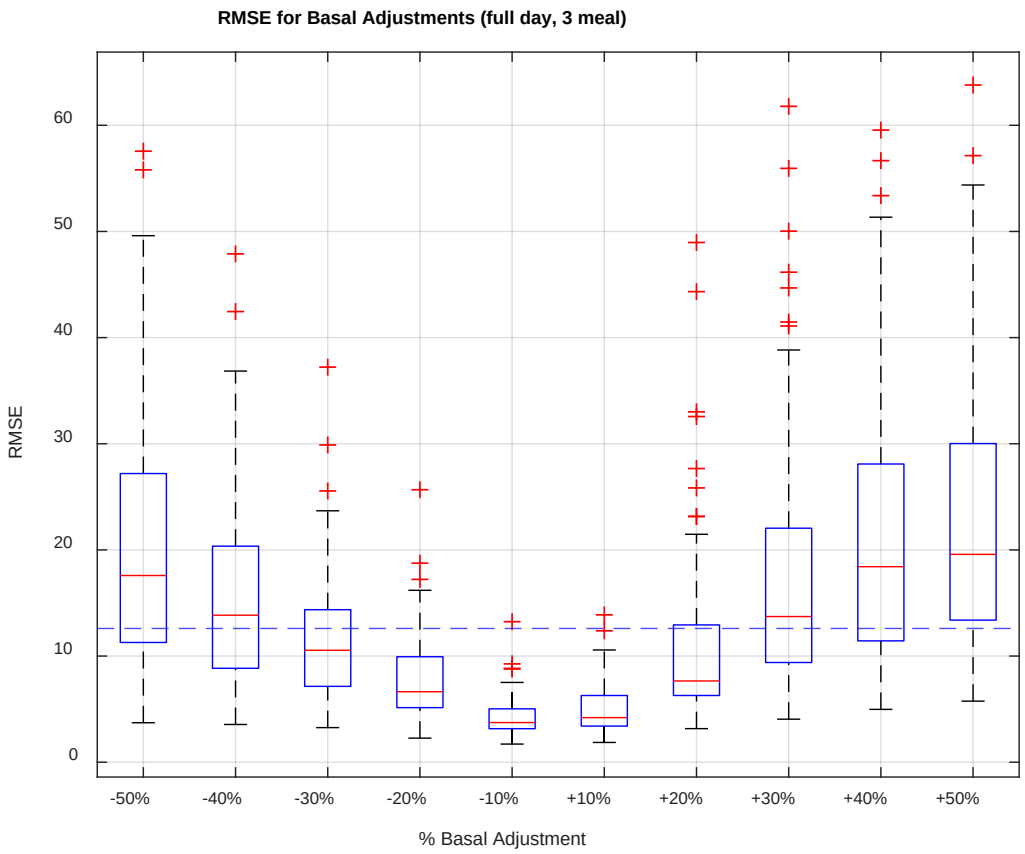


Figure 4.6: Full day RMSE results for net-efect resimulation of basal adjustments for 3 meals.

No zone C or E performance was observed in any of the experiments. For simulation experiments 2 and 3, 100% zone A or B performance (>95% zone A in experiment 2, >85% Zone A in experiment 3) was observed for all experimental scenarios except for the 30% increase in meal bolus which in 1.522% zone D performance. For simulation experiment 4, the basal insulin adjustments, zone A or B performances was greater than 95% up to a 120% increase, with greater than 10% time spent in zone D for basal increases greater than or equal to 130% of the original. Full tables of Clarke error-grid performance are included in the supplementary material located in the appendix.

4.4. Discussion

The above results demonstrate the ability of the updated net effect simulation procedure to faithfully reconstruct the results obtained from the same scenarios replayed in the UVa/Padova simulator for some reasonable adjustments in insulin therapy and meals across the simulated population. Smaller alterations in insulin or meals resulted in smaller errors in the BG signals as reconstructed by the net effect simulation. For meal alterations, total meal removals, and bolus adjustments, the RMSE across the simulator subjects were low through a wide range of scenarios. In cases of "large" meals (180-200% of the original 0.7g/kg BW, or about 100g CHO on average), in both adjustment and removal, performance was somewhat degraded, but still respectable for a large portion of the simulator population. As a comparison, Gani et al [99] considered errors in 60-minute prediction of CGM corresponding to 12.6 mg/dl clinically acceptable (threshold indicated by blue checked line on figs.4.3-4.6). These results give a better idea of the performance of the simulation procedure in a clinical context. For the case of meal alterations and related boluses, the results were encouraging, with across the board A-B range performance greater than 95%. Outcomes from basal adjustments within 20% of the original rate were reproduced accurately; total time in A or B range was greater than 95% on average across all subjects in each scenario (reduction or increase) until at least a 30% increase was given. Outside of this range, the resimulation procedure was less effective in reproducing the reference signals. For reductions in basal rate greater than 20%, Clark error grid performance was still respectable, with time outside of A-B range averaging less than 1%; however, RMSE performance degraded, with median performance above 12.6mg/dl for reductions of 40% and 50%. For the case of basal increases, any simulated increase in basal insulin resulted in some time outside of the A-B range, and, as in the case of reductions, for increases exceeding 20%, the median RMSE exceeded 12.6mg/dl

(though narrowly for the case of 30% increases). This inability to reproduce relatively large changes in basal insulin is a limitation of the net effect simulation procedure to effectively recover the UVa/Padova simulator's nonlinear dynamics in these ranges. Specifically, we note that components of the UVa/Padova simulator, e.g. the nonlinear risk gains active when BG is below its basal value [100], are elements which are unlikely to be accurately reproducible by the net effect simulator's combination of a linear model and additive disturbance. However, we note that such large changes in basal insulin titration are not generally issued in clinical practice—e.g. in [101] basal-rate profile adjustments were algorithmically limited to 25% maximum for *any* segment of the day, let alone total overall adjustments.

For these experiments, the use of simulated data was required in order to secure a "ground truth" comparison to assess the potential domain of validity for the net effect simulation. Unfortunately, this comparator is in itself somewhat flawed, following the best of class but imperfect dynamics of the UVA/Padova simulator. Other models (such as the ones presented by Sorensen [102] or Hovorka et al. [103]) may lead to slightly different results, but none are expected to fully represent actual human derived data. Notwithstanding this fact, the ultimate purpose of the net effect simulation procedure is to allow for simulated modifications of real T1D patients' data by clinicians and engineers in order to evaluate potential changes in therapy. To demonstrate this, we present a "walkthrough" of the application of the net effect simulation using real-world clinical data from a study conducted at the University of Virginia Center for Diabetes Technology.

Given below in fig. 4.7 is an example of field collected data from of our clinical trials (#NCT03394352) and the progression of the above describe simulation procedure.

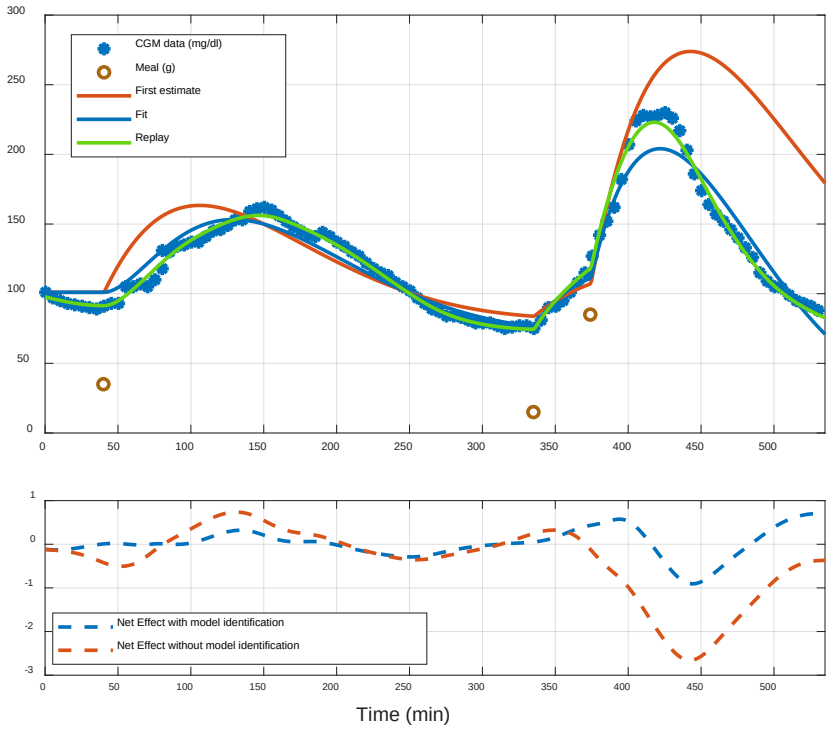


Figure 4.7: Above, plots of BG traces obtained based on real – world clinical CGM data and insulin/meal records. The red trace above shows the open-loop performance of the triangular SOGMM model when given the same insulin meal records as the target data. The blue trace also shows open-loop performance, but after the chosen parameters had been identified on the signal. The green trace shows the fit obtained by both using the identified model and the net effect signal. Below, the dotted traces show the net effect signals for both the population (red) and identified (blue) models.

As demonstrated, the population parameterized triangular SOGMM model does not accurately follow the observed behavior of the subject's CGM data, presenting wide discrepancies, especially following meal intake. There are two sources of this discrepancy—failure of the model's population based parameters to accurately fit the individual subject's meal dynamics and insulin

sensitivity, and the residual variation due to other sources. The purpose of the net effect signal is to account for the latter, while the prior model system identifications step is implemented to account for the former. The fit achieved by first identifying the chosen parameters on the signal and then replying the fitted SOGMM in open-loop based on the data is significantly better, as can be seen both by inspection of the mitigated discrepancies between the CGM data and the CGM values, as well as the smaller deviations from the net effect signal obtained when the deconvolution is implemented using this fitted model (viz. the lower subplot in fig 4.7). When the net effect signal is fed into the identified model, the resulting trace accurately reconstructs the original CGM data.

Using the net effect signal, we can alter the system inputs in order to examine the potential consequences of a change to treatment, specifically its impact on clinically relevant events such as the occurrence of hypoglycemia or hyperglycemia. In the case of the above subject, a carbohydrate treatment was given when the clinician observed that the CGM values dropped below 80mg/dl (minute 345 in Fig 4.7). Partially as a result of this treatment, the subject experienced elevated post-prandial BG readings following the approximately 90 g carbohydrate meal given at minute 375, with some CGM measurements approaching 250 mg/dl. A natural question which follows these observations is what kind of treatment intervention or alteration could have prevented both the hypotreatment and mitigated the related elevation in post-prandial BG level? A candidate option would be an alteration in the insulin dose given with the meal occurring at minute 45. The purpose of the net effect procedure is to simulate just such an intervention (with both alteration of a bolus, at min 45, and the elimination of one occurrence of carbohydrate ingestion, the hypoglycemia treatment). In fig. 4.8, the simulation results of some potential alterations to the

insulin bolus given with the meal occurring at minute 45 of observation, ranging from a 40% decrease to a 40% increase, are given.

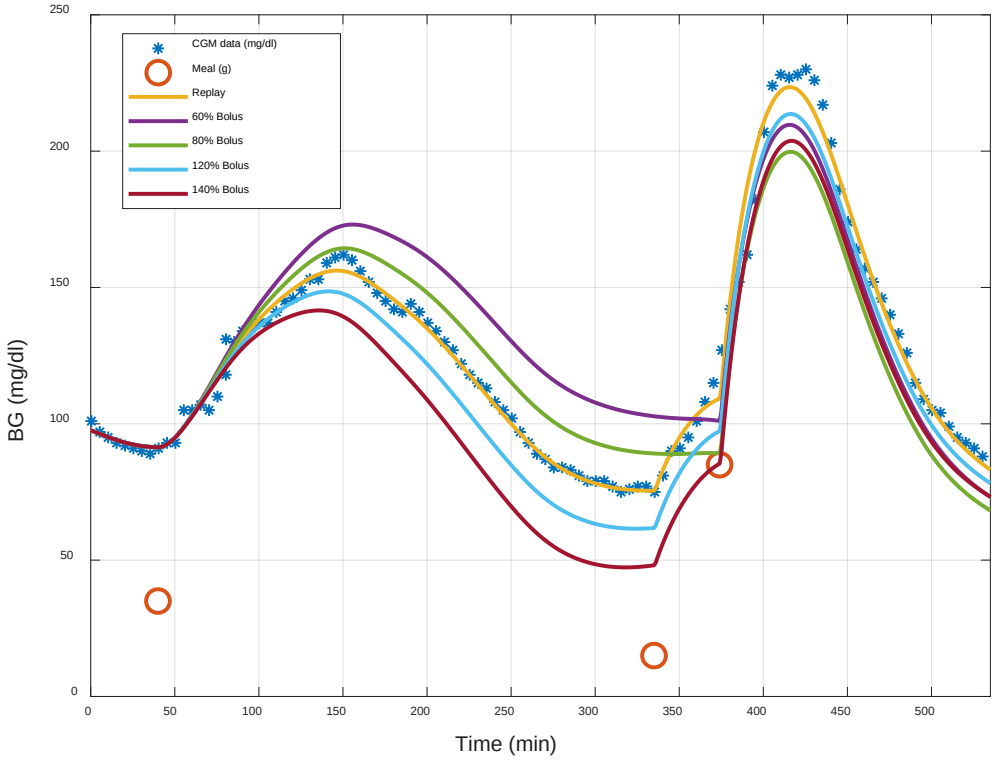


Figure 4.8: BG traces of simulated alterations in the insulin bolus associated with the meal given at minute 45 of observation, together with the CGM signal and meals.

In this case, the net effect simulation indicates that hypotreatment will be avoided if the first meal's insulin bolus is reduced by 20-40%. If the bolus given with the first meal is reduced to 80% of the original, we achieve a case where both the hypotreatment is avoided (BG does not drop below 90mg/dl) and the post-prandial hyperglycemia is significantly reduced (fewer than 5 minutes spent above 180mg/dl). We propose that the use of such a method could powerfully inform clinicians, engineers, and potentially patients themselves when assessing treatment and control regimens for people with T1D. E.g. in situations similar to the above example, a persistent similar

pattern in the subject's data and outcomes of net effect resimulations may indicate that an alteration (here a 20% decrease) in the insulin to carbohydrate ratio is warranted. Simulation via the net effect methodology can help inform clinicians of potential beneficial directions and magnitudes of such changes using the patient's own data, and give engineers an additional method for effectively simulating the T1D glucose-insulin dynamics in order to more efficiently evaluate new technologies within the context of both traditional simulations and clinical trials.

To note, and as mentioned above, the net effect simulation procedure is based on already obtained data, and is meant to reconstruct hypothetical scenarios such as alterations in insulin and meal delivery. Thus we were limited to *in silico* experiments where the ground-truth of the hypothetical scenarios could be recovered. Additionally, while we were able to use the experimental set-up to evaluate the performance of the prior system identification in being able to accurately personalize the SOGMM model, the homogenous nature of the simulated meal dynamics and the lack of variability in insulin sensitivity or other exogenous factors within the chosen simulation scenarios limited the generalizability of the method validation.

4.5. Conclusion

The workflow for the development of new therapies for people with T1D involves many steps, culminating in rigorous clinical trials and, if successful, real world deployment. Patek et. al [4] proposed a simulation methodology for evaluating new insulin therapies based on real world data which we have extended to directly include the capability to evaluate adjustments in meal inputs as well. Additionally, we have updated the underlying SOGMM model, implementing triangular insulin and meal subsystems as well as a prior meal and subject specific system identification procedure in order to personalize the model dynamics to achieve a more accurate recreation of the original data. This updated net effect simulation procedure was able to satisfactorily reproduce

differing scenarios in the UVa/Padova T1D simulator in both clinical and signal estimation metrics. While the ground truth of the hypothetical scenarios which the net effect method is designed to simulate are unrecoverable in real life, the addition of other sources of BG variability, e.g. physical activity or the time varying insulin sensitivity, may allow for more robust evaluations of performance. In the following chapter, this simulation procedure will be combined with the GMAdapt methodology for rapidly producing personalized models for treatment in T1D in order to address potential confounding of the data due to the introduction of systematic interventions.

Chapter 5

5.1 Introduction

In cases where GMAdapt is applied to personalize an advisory/treatment system such as described in chapter 3, the system will be training itself on data generated by a process with which it is interacting, possibly leading to the corruption of the data from the point of view of the learning algorithm. This is a phenomena that has been touched on in recent literature discuss the influence of introducing confounding medical interventions (CMIs) in data used for statistical learning systems [104]. Preventative actions, such as carbohydrate treatments, taken by the user based on the suggestion of the system, will, if successful, eliminate the predicted hypoglycemia, both in actuality and from the data records used for online training. While the elimination of the predicted hypoglycemia in actual practice is the goal of the system, its elimination from the data records introduces *label noise* into the training data [105]. The expression of the features of the data which allowed for the hypoglycemia to be accurately predicted and prevented will now be associated with outcomes where no hypoglycemia occurred. From the perspective of a supervised learning algorithm, this observation will be treated as a false positive, despite the fact that in actuality the prediction and treatment were both appropriate and effective.

There are many questions that can be raised in situations like these. For instance, classification techniques such as the logistic regression model as was used in chapter 3 are known to be robust to cases of *uniform* label noise when trained using standard methods in a static, offline setting [105]. However, their vulnerability to *systematic* label noise such as discussed in the above online setting is, to our knowledge, an unexplored issue. Additionally, there is the question of what is to be done if the introduction of label noise proves to be detrimental to online personalization algorithms such as GMAdapt if it is left unaccounted for.

In the following sections, we will present and evaluate a methodology for overcoming this issue using an *in silico* experiment conducted with the UVa/Padova type 1 diabetes simulator [100] with a new module which simulates the impact of moderate bouts of exercise on glucose-insulin dynamics. This simulation setting provides a unique environment which allows for alternative counter-factual scenarios both with and without preventative carbohydrate treatments (the primary CMIs which affect the data labels in this case) to be observed with all other relevant factors being held equal, and to perfectly recover ground truth labels in the context of an online learning/treatment system such as GMAdapt when applied to the task of hypoglycemia forecasting and prevention.

Five different techniques for addressing the issue of CMIs interfering with the learning process will be investigated:

1. *Do nothing.* Simply ignore the CMIs and see how they impact the online learning procedure if left unaccounted for, relying on the general robustness of logistic regression to label noise in other settings.
2. *Do nothing on treatments.* Adaptation will proceed regularly when no treatments are given, as in technique 1. When treatments are given the algorithm will treat the data as corrupt and no adaptation steps will take place. I.e. we exclude CMIs influenced observations form the training.
3. *Down weigh treatment observations.* As in 1 and 2, when no treatments occur the data the algorithm will proceed as normal. When treatments are given, the observation will be down weighed in accordance with the average likelihood that the treatments prevent hypoglycemia (in the simulation experiment by 81.5%).

4. *Introduce the CMI treatments as a model feature.* In the below case we will be using a logistic regression with hand engineered features as our base forecasting model. A proposed solution for accounting for the CMIs is to treat them as any other confounding variable in a regression setting and introduce them as features by using indicator dummy variables.
5. *Remove the label noise in the data using retrodictive simulations to relabel.* The most sophisticated proposed approach, we use a linear time invariant model deconvolution to generate a net effect signal based on the work of Patek et al [4] as extended in chapter 4 above, allowing for counter-factual simulations derived from the observed data to be generated. We use this procedure to *retrodict* appropriate labels with CMIs removed for training.

In the sections below, we will present methods used to develop personalized hypoglycemia forecasting systems associated with exercise in an online learning setting, and evaluate the resulting impact of giving preventative carbohydrates CMIs on the learning procedure under the five proposed techniques listed above. In the discussion we will address these results and some of the questions that they raise, and the broader issues of CMIs in the context of data driven diabetes treatment systems.

5.2 Methods

5.2.1 Hypoglycemia Risk Forecasting using Logistic Regression

We will use the same basic exercise related hypoglycemia forecasting model as given in chapter 3 above, namely

$$\log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot CGM_{end} + \beta_2 \cdot CGM_{slope} + \beta_3 \cdot IOB_6 \quad (5.1)$$

where π is the probability that a hypoglycemia will occur within two hours of the system query at the commencement of a bout of exercise, CGM_{end} is the final CGM measurement available at the time of the query, CGM_{slope} is the average trend of the CGM trace in the last hour prior to the query, and IOB_6 is the insulin-on-board measure as determined by an exponentially decaying weight of insulin delivered in the previous six hours.

Observed associations of feature variables and responses will take the form $\{x_i, y_i\}$ where $x_i = [1, CGM_{end_i}, CGM_{slope_i}, IOB_{6_i}]$ and $y_i \in \{0,1\}$ will be an indicator variable for whether or not the BG drops below 70mg/dl within the two hours following the exercise related query trigger.

5.2.2 Implementation of GMAdapt for simulated exercise related hypoglycemia prediction and prevention task

In order to personalized the hypoglycemia forecasting logistic regression presented above we implement GMAdapt using constant step sized stochastic gradient descent performed online as new data are generated based on system queries. The purpose of the experiment presented is to assess the impact of CMIs due to carbohydrate ingestion treatments given to prevent predicted hypoglycemia events following exercise on the online learning algorithm. To accomplish this, we will evaluate the five different methods of handling the CMIs as discussed in the introduction above CMIs.

The first method is a straightforward application of GMAdapt updating procedure given in equation chapter 3, and the second two methods, excluding observations in the presence of treatment or down weighing them respectively, require only minor modifications to the algorithm. The fourth method requires a more substantial modification. In this case we will append to the feature vector a variable, "$treat$" indicating whether or not a treatment was given by the system by taking on binary values $treat \in \{0,1\}$, so that $x_i = [1, CGM_{end_i}, CGM_{slope_i}, IOB_{6_i}, treat]$, and extending the parameter vector by an additional term, $\beta = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^T$. With this appended information, the algorithm proceeds following the prediction rule given in equation (5.1).

The fifth method uses the deconvolution procedure to generate an additive net effect signal which allows for the observed scenario to be recreated in retrospective simulations using the methods presented in chapter 4 above. Within these simulations, insulin injections and meal inputs can be altered and the resulting retrodicted traces can be used to relabel the data to account for the impact of the treatment on the outcome.

Using this method, the proposed updating procedure becomes

$$\beta \longleftarrow \beta - \eta(\hat{\pi}_i - \pi_i^-)x_i^T \tag{5.2}$$

where

$$\pi_i^- = \Phi((\delta - \mu_i)/\sigma) \tag{5.3}$$

is the normal cumulative density function transformation of the minimum observed value of the resimulated CGM trace in the relevant time frame used for assessing the y_i value, and δ and σ are associated offset and scale parameters. The use of this transformed, posterior probability function

as opposed to the discrete label used in (5.2) accounts for the uncertainty introduced by the simulation procedure, both in terms of bias and variance introduced.

In order to choice proper values for the μ_i, σ parameters, we investigate the discrepancy between the net effect resimulations of meal removed results and the true replays of the same scenarios obtained using the UVa/Padova simulator. Just as in any prospective assessment of probability of class membership, there is a tradeoff between types of errors (false positives and false negatives) involved in retrospective assessment of probabilities. Different values for the mean parameters μ_i, σ were assessed ranging from $\mu_i = 50$ to 100, σ = 1 to 20, and with a value of $\mu_i = 80$ and $\sigma = 8$ leading to the best performance. Additionally, these parameters have a clinical plausibility, in that a resimulated measure of BG equal to 80mg/dl would be associated with a 50% probability of true hypoglycemia (BG<70mg/dl), with $\sigma = 8$ corresponding to the an average distance of 10% of the median/mean value. This is in agreement with clinical metrics for acceptable accuracy in CGM sensing such as the MARD [95].

5.2.3 In Silico Experimental Design

The in silico experiments were conducted using the 100 virtual adult subjects from the UVa/Padova type 1 diabetes simulator with an exercise simulation module designed to simulate the impact of moderate exercise on glucose-insulin dynamics. 30 days of data were generated for each subject with a breakfast randomly occurring from 6:30am to 7:30am according to a uniform distribution and a normally distributed carbohydrate content proportional to body weight (BW) and saturated at 0 to exclude negative meals with mean 0.4g/kg-BW and standard deviation 0.15g/kg-BW. Likewise, a lunch occurring between 11:00am and 1:00pm with carbohydrate size (normally distributed with mean 0.7g/kg-BW and standard deviation 0.2g/kg/BW) was given each

day. 30 minutes of moderate intensity exercise was simulated at 03:00pm. No dinner meal was given in order to allow for the full impact of the exercise on potential hypoglycemia risk to be observed.

This data (3000 days from 100 subjects) was used to generate population level coefficients for a logistic regression hypoglycemia forecasting model as described in the subsection above. For each of the subjects, these population parameters were used as initialization point and iteratively updated via the process given in equation (2) as new observations arrive, with the updated parameters used to generate predictions on future observations.

Additionally, 30 identical days were generated for each subject using the same random seeds, but including unbolused carbohydrate treatments (0.3g/kg-BW) given immediately before exercise. To simulate CMIs, these days were substituted in for the original days in the adaptation algorithms whenever the classifier's assessment of hypoglycemic risk exceed the treatment threshold. Choices of treatment threshold varying from 0 to 1 in 0.05 increments were evaluated.

The primary outcomes of interest were online classification performance on the undisturbed (no carbohydrate treatment) data in terms of the area under the ROC-AUC and the classifiers sensitivity at a fixed 10% maximum false positive rate.

The five different approaches to handling the CMIs due to carbohydrate consumption in the context of the GMAdapt updates were evaluated through the course of 30 days of updates/observation: (1) take no action to address the potential confounding, (2) addressing the confounding by excluding observations where treatments are given, (3) down weighing such treatment influenced observations by the probability that the hypo was prevented by the treatment, (4) including the treatments in the model as features via an indicator variable, and (5) using the updated net effect simulator to retrodict the true labels and update the observation with this

information. The performance of the population model and the gradient updates performed with unconfounded data serve as a baseline comparison for the effectiveness of each of these approaches.

5.3 Results

Fig 5.1 below presents the ROC curves obtained by using each of the methods (1) -(3) described above to account for the use of preemptive carbohydrate treatments recommended by the system at 3 different thresholds for triggering treatments.

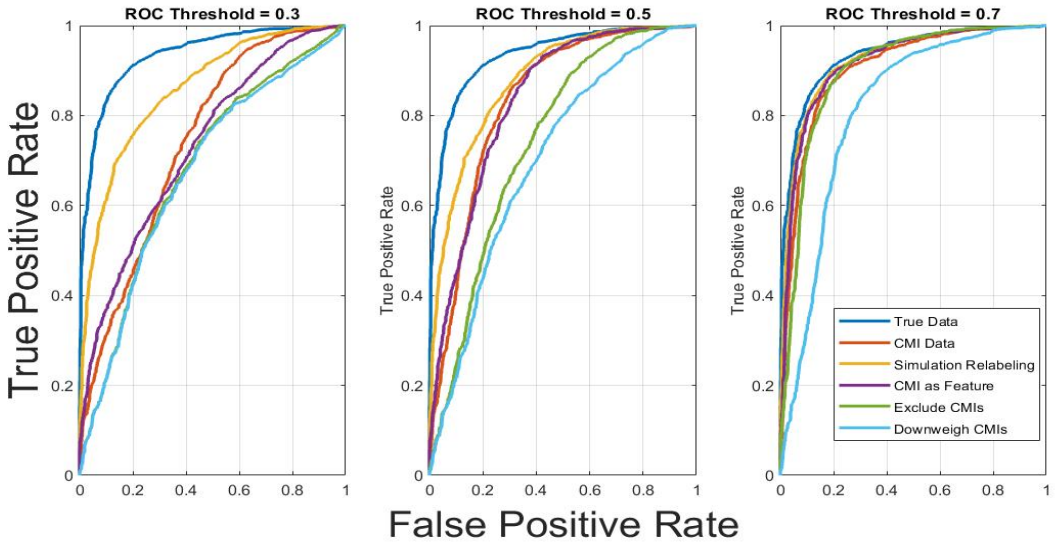


Figure 5.1: ROC plots for each of the 5 tested methods (as well as performances obtained with perfect ground truth knowledge: "true data") at 3 different representative treatment thresholds.

There is a direct trade-off between the obtainable generalizable classification performance and the number of CMIs present in the system. For the three thresholds presenting in Fig 1, the ROC-AUC's obtained by using methods (1) -(5) are given in Table 5.1

Table 5.1

ROC-AUC	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7
(1) No Action	0.7403	0.8343	0.9037
(2)Exclude CMIs	0.6803	0.7432	0.8989
(3) Down weigh CMIs	0.6730	0.7017	0.8117
(4) CMI as Feature	0.7315	0.8350	0.9171
(5) Sim Relabel	0.8581	0.8793	0.9235

Table 5.2 presents the fixed 10% false positive sensitivities for the same scenario

Table 5.2

Fixed 10% FPR Sensitivity	Threshold = 0.3	Threshold = 0.5	Threshold = 0.7
(1) No Action	0.3125	0.4108	0.7320
(2) Exclude CMIs	0.2152	0.2426	0.7217
(3) Down weigh CMIs	0.2142	0.2162	0.3218
(2) CMI as Feature	0.3689	0.4470	0.7957
(3) Sim Relabel	0.6110	0.6291	0.7977

We note that the common interpretation of the ROC curve, that points along the curve represent the tradeoff between sensitivity and specificity, is somewhat inaccurate in the above scenarios. In our scenarios, the choice of decision threshold will itself impact the shape of ROC curves, as seen by the overlaid different plots presented in fig 5.1, somewhat muddling the interpretation. To address this, we present alternative data displays of the tradeoff between the percentage of hypoglycemia able to be detected and prevented by the system, and the performance of the forecasting algorithm in terms of ROC-AUC and Fixed 10% false positive detection rate across varying treatment trigger thresholds, in figs 5.2 and 5.3, respectively. Since the methods of excluding or down weighting of observations with associated CMIs proved to perform more poorly

than doing nothing at all, those methods were excluded from these trade-off analysis plots to improve readability.

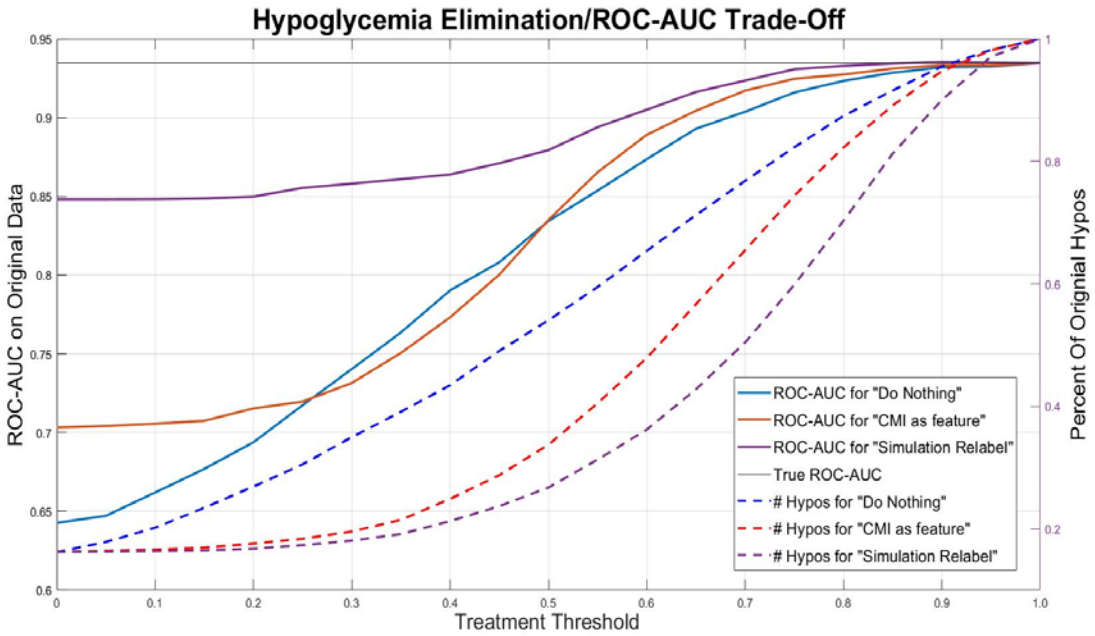


Figure 5.2: Plots representing the number of hypoglycemia events prevented by the system and performance in terms of online ROC-AUC with respect to the ground truth data across different treatment thresholds.

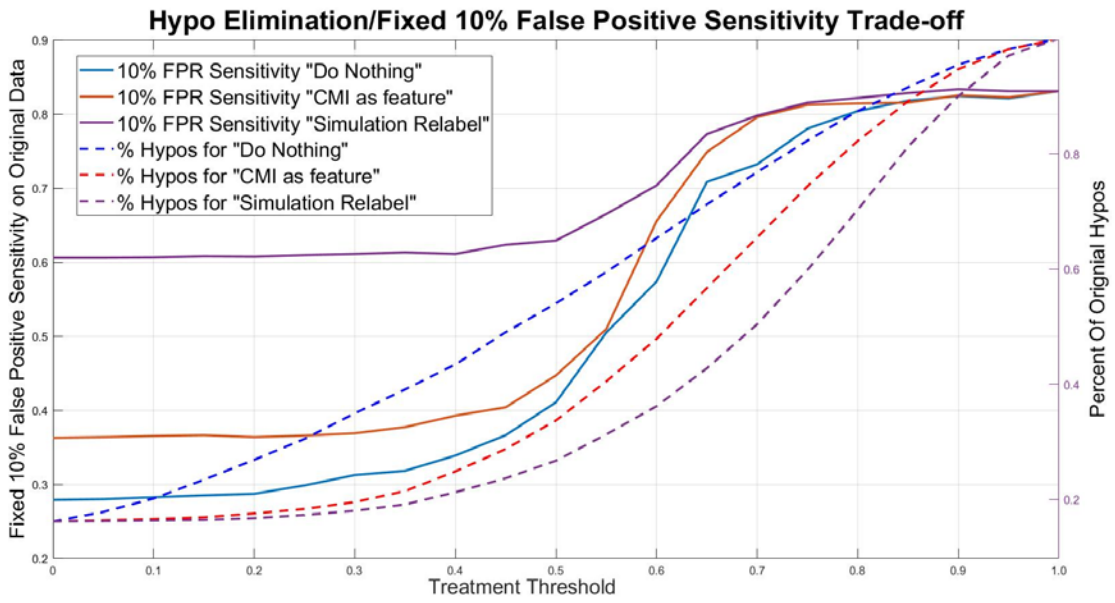


Figure 5.3: Plots representing the number of hypoglycemia events prevented by the system and performance in terms of online Fixed 10 % false positive detection rate with respect to the ground truth data across different treatment thresholds.

5.4 Discussion

The above results show a rather robust performance by the simulation based relabeling approach in terms of its ability to account for the CMIs introduced by carbohydrate treatments. Moving from the relatively high threshold (and thus lower number of CMIs introduced) of 0.7 to the much lower threshold of 0.3 reduced performance in terms of ROC-AUC from 0.9235 to 0.8581, while for the "include treatment as feature" method this reduction went from 0.9171 to 0.7315, in line with the degradation of performance as number of CMIs increases seen in the no action scenario.

These results are qualitatively repeated for the fixed 10% false positive detection rate performance, showing similar drops as number of CMIs increase and domination by the resimulation and relabeling method over the alternative approaches. Of note is the sigmoidal

relationship between treatment threshold and the performance in both metrics as observed in the plotted results in figs 5.3 and 5.4. Fig. 5.3 shows a relative saturation in ROC-AUC performance for the resimulation and relabel method at approximately 0.85 as the threshold, and thus number of hypoglycemic events, drops. Qualitatively similar shapes are shown for the alternatives, though both with steeper slopes and milder saturation, showing a much greater sensitivity to the threshold chosen.

Also of interest in these results is the fact that including treatments as covariates in the model fails to achieve effective parameter estimates over the course of the adaptation relative to the resimulation and relabeling method. At first glance, this is puzzling, since the treatment covariate is causally "downstream" from the system predictions in terms of the impact of the features on the risk of hypoglycemia. Figure 5.4 below shows the proposed causal structure diagram (following Pearl [106]) for the covariates impact on predictions, treatments and hypoglycemia.

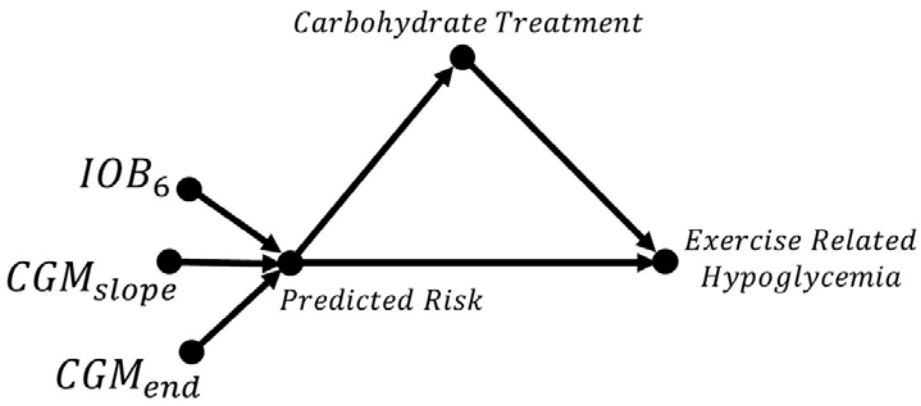


Figure 5.4: Causal diagram of the method 2 approach, indicating carbohydrate treatment is an intervening variable between the predicted risk based on the standard features and the occurrence of exercise related hypoglycemia.

In this case, the carbohydrate treatments are present as intervening, mediating variables, and controlling for them as a feature should allow for the direct effect of predicted risk and its ancillary variables to be properly estimated [106]. In fact, when pseudo-data are generated following the logistic regression modelling assumptions

$$\log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot CGM_{end} + \beta_2 \cdot CGM_{slope} + \beta_3 \cdot IOB_6 + \beta_4 \cdot CHO_{treat} + \epsilon \quad (5.4)$$

With the $[\beta_0, \beta_1, \beta_2, \beta_3]$ coefficients and $[CGM_{end_i}, CGM_{slope_i}, IOB_{6_i}]$ features are drawn according to normal distribution, ϵ represents a Gaussian white noise term, and β_4 = -1.85, chosen so that a in a reduction of the log odds by a factor in agreement with the observed reduction in probability of hypoglycemia obtained in simulation is observed, the correct $[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]$

coefficients are obtained via standard, static model fitting procedures. For example, using 1000 observations generated according to model given by equation (5.4) and the described conditions, with coefficient vector set so that $[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4] = [-0.7914, -0.9272, 1.2363, 0.1653, -1.85]$, the use of Matlab®'s *fitglm* is able to accurately recover satisfactory estimates of all feature coefficients, obtaining $[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4] = [-0.7193, -0.9032, 1.2684, 0.2269, -1.911]$.

The source of the discrepancy between the performance achievable in these numerical simulations and what is observed in the setting presented in our results lies in the convergence properties of online optimization and model fitting methods like constant step-wise stochastic gradient descent. While most methods for fitting generalized linear models in a static, offline setting are asymptotically convergent to the true parameters, and iterate over the data repeatedly, in the online setting each observation is seen only once, and with constant step sizes convergence is only obtained in probability within a neighborhood of the true coefficient values [107], [77] .

To explore this phenomena, we generated 100 trials of 1000 observations using 4 independent, normally distributed random coefficients and features with outcome responses following a logistic regression model as in the simulation above. We implemented a stochastic gradient descent both on this true data, as well as on data influenced by treatments given if the assessed probability was less than the threshold 0.3. "Treatments" in this setting were simulated by a reduction of the true probability by a log odds factor of 1.85 and a separate draw from a Bernoulli distribution with the new reduced probability. The mean results of this numerical exploration across the 100 trials are given in fig. 5.5 below.

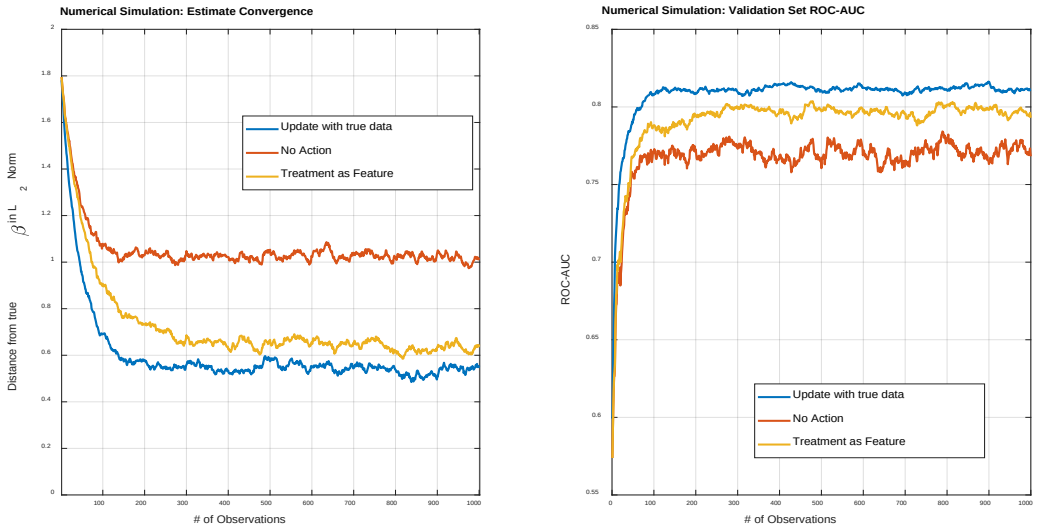


Figure 5.5: Plots representing the empirical, non-asymptotic convergence properties of constant step wise gradient descent under both ground truth labeling CMI induced label noise with and without including the CMIs as covariates.

As can be seen in Fig 5.5, if no action is taken to account for the CMIs, the coefficient estimates stabilize in a neighborhood a significant distance from what is obtained when the true labels are known, leading to a significant and persistent gap in terms of ROC-AUC performance. When treatments are included as features in the model the gap is reduced as more observations accumulate. The counter-intuitive discrepancy in the performance in the online setting—despite the causal structure indicating that controlling for treatments should allow for an estimation of the direct effect of the predictor variables— is in actuality a result of the *non-asymptotic* convergence properties of the corresponding models. As can be seen, the mean distance from the true coefficients in terms of the L_2 norm for both the "no action" and "treatment as feature" approaches lag behind the approach using ground truth data (what would be simulated in the case of our simulation and relabeling approach above). They do not diverge from each other noticeable until

approximately 30-50 observations are seen (fig 5.6), which qualitatively corresponds with the *in silico* results shown. In application, the range of interest will be towards the lower end of the domain in terms of the number of observations.

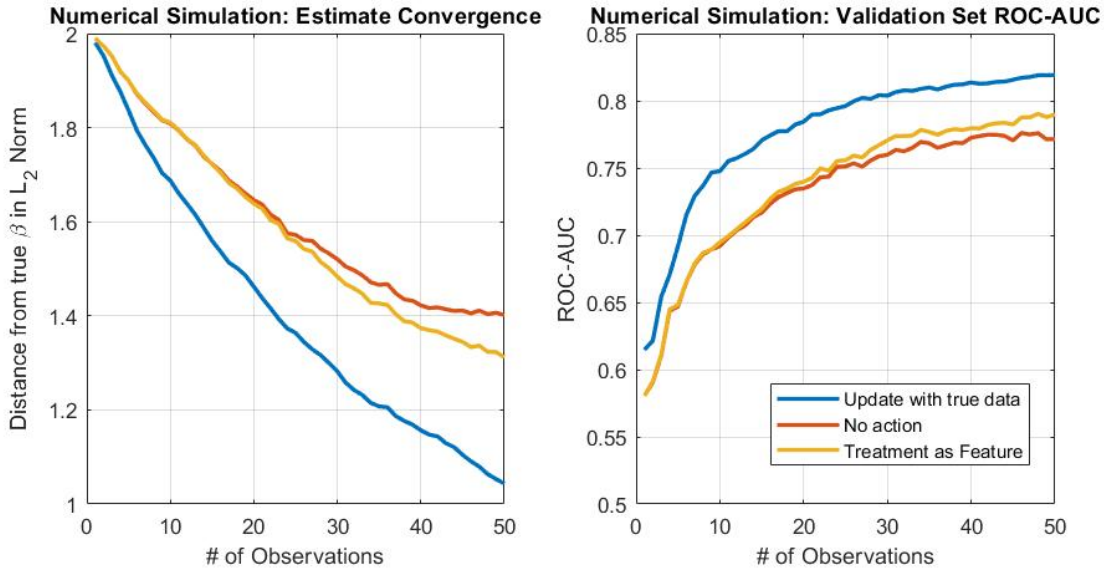


Figure 5.6: Zooms of plots presented in fig 5.5. Focusing on first 50 observations for non-asymptotic analysis

For example, if a person regularly exercises three days a week, then it would be expected that 30 observations would not occur for 90 days. In this case, they ability of a model which includes treatment as a feature to asymptotically capture the direct effect of the other features in spite of the CMI induced label noise is of little value. To the degree to which a resimulation procedure such as used above is able to accurately capture the counter-factual ground truth it will perform significantly better in the domain of interest. The simulation based results above indicate that this net effect resimulation and relabeling procedure is a promising approach for accomplishing this task.

5.5 Conclusion

The above presented in silico results show that the introduction of label noise due to system triggered carbohydrate ingestion presents a hurdle for obtaining online personalization of predictive models by methods such as GMAdapt in the field of T1D. Integrating within the system a deconvolution based retrospective simulation of observed data such as described in chapter 4 above allows for an accurate recovery of approximate ground truth observation labels in the presences of CMI induced label noise. A potential theoretically viable alternative—including these CMIs as feature variables in the model— fails comparatively in practical terms due to the initial non-asymptotic convergence properties given the relevant domain size in terms of available observations for the task.

The effect of CMIs on online learning algorithms such as GMAdapt, both in the context of T1D and in broader medical applications, is a field of research which will need to be thoroughly explored as new technology, available data, and statistical learning techniques expand in the years ahead and the need for effective methods of personalization grows.

Conclusions and contributions

The significant physiological and behavioral heterogeneity present in conditions such T1D, and in medical conditions more generally, necessitates the development of technology able to effectively achieve personalized treatments within the constrained contexts of real-world deployment. Our contribution to this task, presented in this dissertation, is the development of an online optimization procedure design to fit within the kinds of systematic constraints encountered in the implementation of data informed decision support systems for T1D, specifically as applied to the task of hypoglycemia forecasting and prevention in sensitive, risky contexts. The systematic approach taken allows for aggregated data to be leveraged to the task of developing feasible baseline models, which are then rapidly adapted to the unique data streams of individual users, achieving personalized model performance as data accumulates in a scalable, deployable system. The specific contributions of the research presented in this dissertation can be enumerated as

1. The development of a methodology for leveraging population data to apply to tasks which, due to heterogeneity, require personalized or subgroup specific models to be implemented effectively, by means of online learning techniques. The method, GMAdapt, is especially useful for achieving high performance for simpler, intelligible models such as those based on logistic regression, which have benefits in contexts where more sophisticated “black box” methods face significant barriers and constraints to implementation due to data, safety, and other stakeholder driven factors such as user confidence. Once initialized, this methodology allows for the autonomous development of personalized decision rules, and is easily scalable and deployable due to its low complexity and low computational burden. We note that the basic structure: population data used for feature selection and validation, leading to individual level deployment and learning/personalization, is

essentially agnostic to the underlying structure of the model or task, provided that it fits within the context of supervised learning paradigm and is amiable to gradient based optimization methods.

2. An extension of the domain of validity for the net effect deconvolution based resimulation technique originally presented in Patek et. al by means of prior system identification of a modified version of the underlying model introducing “triangular” insulin-glucose dynamics. The identification of the insulin sensitivity, glucose availability, and multithread identification of meal transport parameters allow for more accurate reconstructions of historical signals and attenuates issues of model-mismatch leading to inappropriate net effect signal artifacts, particularly around meals.

This extension of the net effect methodology in itself delivers a valuable contribution to the field. It enables the assessment of hypothetical scenarios such as altered treatments when applied in the context of therapy for diabetes based on a subjects own, real-world data. It can aid in the design and evaluation of experiments for new treatments and therapies, and has potential to inform users and their physicians directly when integrated into larger clinical systems. For example, it can allow users and their physicians themselves to “play” with their data, by simulating different behavioral or treatment regimens and observing the hypothetical reconstructions of those scenarios in simulation, and to choose new behavior or treatment patterns based on this information

3. We integrate the extended net effect methodology together with GMAdapt in order to address the impact of confounding medical interventions for learning systems in the context of type 1 diabetes decision support systems. Confounding interventions present what we consider an underexplored challenge to the effective, real world implementation of supervised learning based

systems, as many of the assumptions which underlay traditional statistical learning approaches to regression or classification tasks are invalidated when actions are taken based on data which affect the label in a manner which is not appropriately compensated for in the modeling process. Methods of addressing this problem by direct deconfounding, which work satisfactorily in the context of static, offline model fitting, perform poorly in the online learning setting, where the multiple passes through the data which are necessary to properly estimate the deconfounded coefficients are not possible. We implemented an alternative approach, which proved amiable to the context of online learning, wherein we used the new net effect simulator to remove confounding interventions and then relabel the data based on the reconstructed signals obtained. Additionally, we used a probabilistic transform of this result which proved especially suitable to the gradient based updating of logistic regression classifiers.

The overall system developed through the course of this dissertation research allows for rapid, autonomous model personalization in the context of decision support systems in type 1 diabetes, even in the presences of interventions which themselves are informed by the systems outputs. The limitations of such a system will largely be determined by the degree to which a potential application fits a structure similar to that observed in the hypoglycemia forecasting task, and the relative performance of alternative methods for achieving personalized or subset specified models. For instance, our explorations indicate that in terms of final metric performance, hierarchical modeling techniques for the development of personalized models in type 1 diabetes perform comparably to GMAdapt in straightforward application. However, the computational burden, scalability, and deployment of systems based on such hierarchical models, which must recalculate coefficients based on the entire available dataset for each iteration, are, in our opinion,

not justified by the meager to statistically nonexistent gains delivered over the computationally and conceptually simple, highly deployable GMAdapt methodology.

We also not that there is a degree of modularity in GMAdapt, which allows for other, suitable modeling procedure to be used. Any supervised learning method which is amiable to a stochastic gradient optimization can be “plugged in”, and the use of logistic regression for the application tasks in this research does not constrain future implementations for the same or other tasks. Support vector machines, neural networks, linear regressions, all can be updated via stochastic gradient methods, and may be suitable to use in the context of a GMAdapt deployment for a specific task, with due consideration being given to relevant differences.

Additionally, we reiterate and expand on the fact that, while informed by the constraints in factors involved in diabetes treatment, GMAdapt is not essentially tied down to this field of application. The essential features are the requirement for rapidly bridging the gap between population and personalized models in a context where significant benefits are to be expected from such personalization, but the data available is too sparse to effectively apply traditional, offline, methods to achieve individualized models.

As an example application outside the field of diabetes, consider a hypothetical case in electrical engineering of online load prediction for a new gas turbine power plant. There is likely to be significant heterogeneity from location to location in the true relationship between predictive factors and optimal demand schedule for power. These differences can be due to climate, variations in seasonality, culture, geography, etc. and it may not be discernable how these factors will come together and effect the outcomes in any individual case based on prior knowledge. In this case, it may be suitable to initialize the system using a population level model (or a model based on a subset of the population deemed sufficiently similar to the new plant) and utilize

GMAdapt to update the model as data specific to the new plant becomes available—with simulation methodologies used in a similar manner to remove confounding interventions from the data.

As the application of techniques of machine learning and data science to engineering problems inevitably increase in future years, methods will need to be developed which can apply these powerful techniques while overcoming the constraints face in real-world engineering problems. A key issue will be dealing with the heterogeneity commonly present in application fields when the point of intervention of the system occurs at the level of the varying individuals or subgroups, and for our methods to move beyond the actuarial style calculations of traditional statistical learning by effectively implementing personalized system models. GMAdapt presents the mode of approach that will be necessary to solve the type of systematic problems likely to be encountered as advanced numerical and mathematical methods are combined with data and new technologies to solve real-world engineering problems in the coming years.

References

[1] J. Pearl, "The Seven Tools of Causal Inference with Reflections on Machine Learning," *Communications of the ACM,* vol. 62, no. 3, pp. 54-60, 2019.

[2] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm and N. Elhadad, "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission," in *Knowledge Discovery and Data Mining*, Sydney, Austrailia, 2015.

[3] C. Paxton and S. Saria, "Developing Predictive Models Using Electronic Medical Records: Challenges and Pitfalls," in *AMIA Annual Symposium Preceedings*, Washington, DC, 2013.

[4] S. D. Patek, D. Lv, E. A. Ortiz, C. Hughes-Karvetski, S. Kulkarni, Q. Zhang and M. D. Breton, "Empirical Representation of Blood Glucose Variability in a Compartmental Model," in *Prediction Methods for Blood Glucose Concentration*, Springer, Cham, 2016, pp. 133-157.

[5] C. C. Quianzon and I. Cheikh, "History of insulin," *Journal of Community Hospital Internal Medicine Perspective*, vol. 2, no. 2, pp. 1-3, 2012.

[6] J. A. Todd, "Etiology of Type 1 Diabetes," *Immunity*, vol. 32, pp. 457-467, 2010.

[7] T. L. Van Belle, K. T. Coppieters and M. G. Von Herrath, "Type 1 Diabetes: Etiology, Immunology,and Therapeutic Strategies," *Physiology Review*, vol. 91, pp. 79-118, 2011.

[8] M. A. Atkinson, G. S. Eisenbarth and A. W. Michels, "Type 1 diabetes," *The Lancet*, vol. 383, no. 9911, pp. 69-82, 2014.

[9] L. Kahanovitz, P. M. Sluss and S. J. Russell, "Type 1 Diabetes -- A Clinical Perspective," *Point Care,* vol. 16, no. 1, pp. 37-40, 2017.

[10] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care,* vol. 34, no. Supplement 1, pp. S81-S90, 2014.

[11] R. J. McCrimmon and R. S. Sherwin, "Hypoglycemia in Type 1 Diabetes," *Diabetes,* vol. 59, no. 10, p. 2333–2339, 2010.

[12] G. Shafiee, M. Mohajeri-Tehrani, M. Pajouhi and B. Larijani, "The importance of hypoglycemia in diabetic patients," *Journal of Diabetes and Metabolic Disorders,* vol. 11, no. 17, 2012.

[13] A.-S. Brazeau, R. Rabasa-Lhoret, I. Strychar and H. Mircescu, "Barriers to physical activity among patients with type 1 diabetes," *Diabetes Care,* vol. 31, no. 11, pp. 2108-2109, 2008.

[14] S. R. Colberg, R. Laan, E. Dassau and K. David, "Physical Activity and Type 1 Diabetes: Time for a rewire?," *Journal of Diabetes Science and Technology,* vol. 9, no. 3, pp. 609-618, 2015.

[15] F. Porcellati, P. Lucidi, G. B. Bolli and C. G. Fanelli, "Thirty Years of Research on the Dawn Phenomenon: Lessons to Optimize Blood Glucose Control in Diabetes," *Diabetes Care,* vol. 36, no. 12, pp. 3860-3862, 2013.

[16] K. Trout, M. Rickels, M. Schutta, M. Petrova, E. Freeman, N. Tkacs and K. Teff, "Menstrual cycle effects on insulin sensitivity in women with type 1 diabetes: a pilot study.," *Diabetes Technology and Therapeutics,* vol. 9, no. 2, pp. 176-82, 2007.

[17] D. Dabelea, "The accelerating epidemic of childhood diabetes," *The Lancet,* vol. 373, no. 9680, pp. 1999-2000, 2009.

[18] D. M. Maahs, N. A. West, L. J. M and M.-D. E. J, "Epidemiology of Type 1 Diabetes," *Endocrinology and Metabolism Clinics,* vol. 39, no. 3, pp. 481-497, 2010.

[19] K. Bullard, C. Cowie, S. Lessem, S. Saydah, A. Menke, L. Geiss, T. Orchard, R. DB and G. Imperatore, "Prevalence of Diagnosed Diabetes in Adults by Diabetes Type — United States, 2016," *Morbidity and Mortality Weekly Report,* vol. 67, no. 12, p. 359–361, 2018 .

[20] M. Bliss, "The History of Insulin," *Diabetes Care,* vol. Supp, no. 3, pp. 4-7, 1993.

[21] "The Nobel Prize in Physiology or Medicine 1923," Nobel Media AB , 2019. [Online]. Available: <https://www.nobelprize.org/prizes/medicine/1923/summary/> . [Accessed 1 October 2019].

[22] The 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 New England Journal of Medicine,* vol. 329, no. 14, pp. 977-86, 1993.

[23] B. Paldus, M. H. Lee and D. N. O'Neal, "Insulin pumps in general practice," *Australia Prescriber,* vol. 41, no. 6, p. 186–190, 2018.

[24] "Glycaemic control with continuous subcutaneous insulin infusion compared with intensive insulin injections in patients with type 1 diabetes: meta-analysis of randomised controlled trials.," *BMJ,* vol. 324, 2002.

[25] S. Clarke and J. Foster, "A history of blood glucose meters and their role in self-monitoring of diabetes mellitus," *British Journal of Biomedical Science,* vol. 69, no. 2, pp. 83-93, 2012.

[26] S. SI, H. Khan, A. Ekhzaim, A. Masood and M. Sakharkar, "Significance of HbA1c Test in Diagnosis and Prognosis of Diabetic Patients," *Biomarker Insights,* vol. 11, pp. 95-104, 2016.

[27] D. Rodbard, "Continuous Glucose Monitoring: A Review of Successes, Challenges, and Opportunities," *Diabetes Technology and Therapeutics,* vol. 18, no. Suppl 2, pp. S2-3–S2-13, 2016.

[28] D. Rodbard, "Continuous Glucose Monitoring: A Review of Recent Studies Demonstrating Improved Glycemic Outcomes," *Diabetes Technology and Therapeutics ,* vol. 19, no. Suppl 3, pp. S-25–S-37, 2017.

[29] C. Cobelli, C. Dalla Man, G. Sparacino, L. Magni, G. De Nicolao and B. P. Kovatchev, "Diabetes: Models, Signals, and Control," *IEEE Review of BioMedical Engineering,* vol. 2, no. 1, pp. 54-96, 2009.

[30] E. Dassau, E. Atlas and M. Phillip, "Closing the loop," *The International Journal of Clinical Practice Supplement,* vol. 170, pp. 20-25, 2011.

[31] A. Clemons, P. Chang and R. Myers, "The development of Biostator, a Glucose Controlled Insulin Infusion System," *Hormone and Metabolic Research,* vol. Supplement 7, pp. 23-33, 1977.

[32] C. Cobelli, E. Renard and K. Boris, "Artificial Pancreas: Past, Present, Future," *Diabetes,* vol. 60, no. 11, p. 2672–2682, 2011.

[33] B. Kovatchev, W. V. Tamborlane, W. T. Cefalu and C. Cobelli, "The Artificial Pancreas in 2016: A Digital Treatment Ecosystem for Diabetes," *Diabetes Care,* vol. 39, no. 7, pp. 1123-1126, 2016.

[34] R. Nimri, A. R. Ochs, J. E. Pinsker, M. Phillip and E. Dassau, "Decision Support Systems and Closed Loop," *Diabetes Technology and Therapeutics,* vol. 21, no. Supplement 1, pp. S42-S56, 2019.

[35] R. N. Bergman, Y. Z. Ider, C. R. Bowden and C. Cobelli, "Quantitative Estimation of insulin sensitivity," *American Journal of Physiology Endocrinology and Metabolism,* vol. 236, no. 6, pp. E667-77, 1979.

[36] R. N. Bergman, "Minimal Model: Perspective from 2005," *Hormone Research,* vol. 64, no. suppl 3, pp. 8-15, 2005.

[37] B. Kovatchev, M. Breton, C. Dalla Man and C. Cobelli, "In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes.," *Journal of Diabetes Science and Technology,* vol. 3, no. 1, pp. 44-55, 2009.

[38] C. Dalla Man, R. A. Rizza and C. Cobelli, "Meal Simulation Model of the Glucose-Insulin System," *IEEE Transactions in Biomedical Engineering,* vol. 54, no. 10, pp. 1740-1740, 2007.

[39] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev and C. Cobelli, "The UVA/PADOVA Type 1 Diabetes Simulator New Features," *Journal of Diabetes Science and Technology,* vol. 8, no. 1, pp. 26-34, 2014.

[40] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, B. M, B. Kovatchev, C. Dalla Man and C. Cobelli, "The UVA/Padova Type 1 Diabetes Simulator Goes From

Single Meal to Single Day," *Journal of Diabetes Science and Technology,* vol. 12, no. 2, p. 273–281, 2018.

[41] C. Dalla Man, A. Caumo and C. Cobelli, "The oral glucose minimal model: estimation of insulin sensitivity from a meal test.," *IEEE Transactions in Biomedial Engineering,* vol. 49, no. 5, pp. 419-29, 2002.

[42] R. N. Bergman, Y. Z. Ider, C. R. Bowden and C. Cobelli, "Quantitative estimation of insulin sensitivity.," *The American journal of physiology,* vol. 236, no. 6, pp. E667-E677, 1979.

[43] C. D. Man, A. Caumo, R. Basu, R. Rizza, G. Toffolo and C. Cobelli, "Minimal model estimation of glucose absorption and insulin sensitivity from oral test: validation with a tracer method," *American Journal of Physiology Endocrinology Metabolism* , vol. 287, no. 4, pp. E637-E643, 2004.

[44] B. Kovatchev, S. Patek, E. Ortiz and M. Breton, "Assessing sensor accuracy for non-adjunct use of continuous glucose monitoring," *Diabetes Technology and Therapeutics,* vol. 17, no. 3, pp. 177-186, 2015.

[45] B. Ozaslan, "Personalized Optimization of Insulin Treatment Policies for the Management of Physical Activity in Patients with Type 1 Diabetes," University of Virginia, Department of Systems Engineering, PHD (Doctor of Philosophy) dissertation, Charlottesville, Virginia, 2019.

[46] M. Vettoretti, A. Facchinetti, G. Sparacino and C. Cobelli, "Predicting Insulin Treatment Scenarioswith the Net Effect Method: Domain of Validity," *Diabetes Technology and Therapeutics,* vol. 18, no. 11, pp. 694-704, 2016.

[47] J. Pearl, "The seven tools of causal inference with reflectionson machine learning," Communications of Association for Computing Machinery, 2018.

[48] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

[49] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, New York: Springer, 2009.

[50] G. P. Forlenza, Z. Li, B. A. Buckingham, J. E. Pinsker, E. Cengiz, R. P. Wadwa, L. Ekhlaspour, M. M. Church, S. A. Weinzimer, E. Jost, T. Marcal, C. Andre, L. Carria, V. Swanson and J. W. Lum, "Predictive Low-Glucose Suspend Reduces Hypoglycemia in Adults, Adolescents, and Children With Type 1 Diabetes in an At-Home Randomized Crossover Study: Results of the PROLOG Trial," *Diabetes Care* , vol. 41, no. 10, pp. 2155-2161, 2018.

[51] C. Zecchin, A. Facchinetti, G. Sparacino and C. Cobelli, "How Much Is Short-Term Glucose Prediction in Type 1 Diabetes Improved by Adding Insulin Delivery and Meal Content Information to CGM Data? A Proof-of-Concept Study," *Journal of Diabetes Science and Technology*, vol. 10, no. 5, p. 1149–1160, 2016.

[52] M. Eren-Oruklu, A. Cinar, L. Quinn and D. Smith, "Estimation of Future Glucose Concentrations with Subject-Specific Recursive Linear Models," *Diabetes Technology and Therapeutics*, vol. 11, no. 4, pp. 243-253, 2009.

[53] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti and C. Cobelli, "Glucose Concentration can be Predicted Ahead in Time From Continuous Glucose Monitoring Sensor Time-Series," *IEEE Transactions in Biomedical Engineering*, vol. 54, no. 5, pp. 932-937, 2007.

[54] P. McCullagh and J. Nelder, Generalized Linear Models, London: Chapman & Hall, 1989.

[55] J. A. Nelder and R. Wedderburn, "Generalized LInear Models," *Journal of the Royal Statistical Society,* vol. 135, no. 3, pp. 370-384, 1972.

[56] T. P. Minka, "A comparison of numerical optimizers for logistic regression," 26 March 2007. [Online]. Available: https://pdfs.semanticscholar.org/69da/9d1059d91c3536067330085f6a2233592a39.pdf. [Accessed 6 October 2019].

[57] A. Agresti, Categorical Data Analysis, Hoboken, NJ: Wiley, 2014.

[58] W. W. Stroup, Generalized Linear Mixed Models Modern Concepts, Methods and Applications, Boca Raton: Chapman & Hall, 2013.

[59] S. M. Butler and T. A. Louis, "Consistency of Maximum Likelihood Estimators In General Random Effects Models for Binary Data," *The Annals of Statistics,* vol. 25, no. 1, pp. 351-377, 1997.

[60] A. Gelman, "Analysis of variance--why it is more important than ever," *The Annals of Statistics,* vol. 33, no. 1, pp. 1-33, 2005.

[61] C. E. McCulloch and S. R. Searle, Generalized, Linear, and Mixed Models, Hoboken, NJ: Wiley-Interscience, 2004.

[62] A. Gelman and J. Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge, UK: Cambridge University Press, 2007.

[63] C. E. McCulloch and J. M. Neuhaus, "Prediction of random effects in linear and generalized linear models under model misspecification," *Biometrics,* vol. 67, no. 1, pp. 270-279, 2011.

[64] G. Robinson, "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science,* vol. 6, no. 1, pp. 15-51, 1991.

[65] S. L. Zeger and M. R. Karim, "Generalized Linear Models With Random Effects: A Gibbs Sampling Approach," *Journal of the American Statistical Association,* vol. 86, no. 413, pp. 79-86, 1991.

[66] D. Gamerman and H. F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Boca Raton: Chapman & Hall, 2006.

[67] V. Vapnik, "Principles of Risk Minimization for Learning Theory," in *Advances in neural information processing systems*, 1992.

[68] V. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Transactions on Neural Networks,* vol. 10, no. 5, pp. 988-999, 1999.

[69] M. S. Pepe, The Statistical Evaluation of Medical Tests for Classification and Prediction, New York: Oxford University Press, 2002.

[70] X.-H. O. N. A. Zhou and D. K. McClish, Statistical Methods in Diagnostic Medicine, Hoboken: Wiley, 2011.

[71] P. C. Austin and E. W. Steyerberg, "Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable," *BMC Medical Research Methodology*, vol. 12, no. 82, 2012.

[72] E. R. Delong, D. M. Delong and D. L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Reciever Operating Characteristic Curves: A Nonparametric Approach," *Biometrics,* vol. 44, no. 3, pp. 837-845, 1988.

[73] P. Allison, "Convergence failures in logistic regression," in *Proceedings of the SAS Global Forum*, San Antonio, 2008.

[74] A. Gelman, A. Jakulin, M. G. Pittau and Y.-S. Su, "A Weakly Informatie Default Prior Distribution for Logistic and Other Regression Models," *The Annals of Applied Statistics,* vol. 8, no. 4, pp. 1360-1383, 2008.

[75] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[76] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, Paris France, 2010.

[77] L. Bottou, F. E. Curtis and J. Nocedal, "Optimization Methods forLarge-Scale Machine Learning," *SIAM Review,* vol. 60, no. 2, pp. 223-311, 2018.

[78] F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," *The New England Journal of Medicine,* no. 372, pp. 793-795, 2015.

[79] S. S. Rich and W. T. Cefalu, "The Impact of Precision Medicine in Diabetes: A Multidimensional Perspective," *Diabetes Care,* vol. 39, no. 11, pp. 1854-1857, 2016.

[80] E. Simpson, "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society,* vol. 13, no. 2, pp. 238-241, 1951.

[81] S. Ameringer, R. Serlin and S. Ward, "Simpson's paradox and experimental research," *Nursing research ,* vol. 58, no. 2, pp. 123-127, 2009.

[82] A. Tebani, C. Afonso, S. Marret and S. Bekri, "Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations," *International Journal of Molecular Sciences,* vol. 17, pp. 1555-1582, 2016.

[83] N. S. Abul-Husn and E. Kenny, "Personalized Medicine and the Power of Electronic Health Records," *Cell,* vol. 177, no. 1, pp. 58-69, 2019.

[84] R. Bachran, P. Beyer, C. Klinkert, B. Heidtmann, J. Rosenbauer and R. W. Holl, "Basal rates and circadian profiles in continuous subcutaneous insulin infusion (CSII) differ for preschool children, prepubertal children, adolescents and young adults.," *Basal rates and circadian profiles in continuous subcutaneous insulin infusion (CSII) differ for preschool children, prepubertal children, adolescents and young adults.* , vol. 13, no. 1, pp. 1-5, 2012.

[85] D. Rodbard, "Optimizing the Estimation of Carbohydrate-to-InsulinRatio and Correction Factor," *Diabetes Technology & Therapeutics,* vol. 20, no. 2, pp. 94-97, 2018.

[86] G. B. Bolli, D. Kerr, R. Thomas, E. Torlone, A. Sola-Gazagnes, E. Vitacolonna, J. L. Selam and P. D. Home, "Comparison of a Multiple Daily Insulin Injection Regimen (Basal Once-Daily Glargine Plus Mealtime Lispro) and Continuous Subcutaneous Insulin Infusion (Lispro) in Type 1 Diabetes," *Diabetes Care* , vol. 32, no. 7, pp. 1170-1176. , 2009 .

[87] S. D. Patek, D. Lv, E. Campos-Nanez and M. Breton, "Retrospective Optimization of Daily Insulin Therapy Parameters: Control Subject to a Regenerative Disturbance Process," in *11th IFAC Symposium on Dynamics and Control of Process Systems Including Biosystems DYCOPS-CAB* , Trondheim, Norway, 2016.

[88] C. Kennedy and J. Turley, "Time series analysis as input for clinical predictive modeling: Modeling cardiac arrest in a pediatric ICU," *Theoretical biology & medical modelling,* vol. 8, no. 40, 2011.

[89] L. Bottou, "Large-scale machine learning with stochastic gradient descent," *Proceedings of COMPSTAT'2010,* pp. 177-186, 2010.

[90] M. Breton, S. Patek, D. Lv, E. Schertz, J. Robic, J. Pinnata, L. Kollar, C. Barnett, C. Wakeman, M. Oliveri, C. Fabris, Chernavvsky, K. B. D. and S. Anderson, "Continuous Glucose Monitoring and Insulin Informed Advisory System with Automated Titration and Dosing of Insulin Reduces Glucose Variability in Type 1 Diabetes Mellitus.," *Diabetes Technology & Therapeutics,* vol. 20, no. 8, pp. 531-540, 2018.

[91] P. Peduzzi, J. Concato, E. Kemper, T. Holford and A. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis.," *Journal of Clinical Epidemiology,* vol. 49, no. 12, pp. 1373-9, 1996.

[92] A. Gelman and J. Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge, UK: Cambridge University Press, 2007.

[93] R. Wolfinger and M. O'connell, "Generalized linear mixed models a pseudo-likelihood approach," *Journal of Statistical Computation and Simulation,* vol. 48, no. 3-4, pp. 233-243, 1993.

[94] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, Waltham, MA: Morgan Kaufman Publishers, 2012.

[95] B. P. Kovatchev, S. D. Patek, E. A. Ortiz and M. D. Breton, "Assessing Sensor Accuracy for Non-AdjunctUse of Continuous Glucose Monitoring," *Diabetes Technology & Therapeutics,* vol. 17, no. 3, pp. 177-186, 2015.

[96] M. Vettoretti, A. Facchinetti, G. Sparacino and C. Cobelli, "Predicting Insulin Treatment Scenarioswith the Net Effect Method:Domain of Validity," *Diabetes Technology and Therapeutics,* vol. 18, no. 11, pp. 694-704, 2016.

[97] G. Bellu, M. P. Saccomani, S. Audoly and L. D'Angiò, "DAISY: a new software tool to test global identifiability of biological and physiological systems," *Computer methods and programs in biomedicine,* vol. 88, no. 1, pp. 52-61, 2007.

[98] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter and S. L. Pohl, "Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose," *Diabetes Care,* vol. 10, no. 5, pp. 622-628, 1987.

[99] A. Gani, A. Gribok, S. Rajaraman, W. Ward and J. Reifman, "Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling," *IEEE Transactions in Biomedical Engineering,* vol. 56, no. 2, pp. 246-254, 2009.

[100] R. Visentin, E. Campos-Náñez, M. Schiavon, D. Lv, M. Vettoretti, M. Breton, B. P. Kovatchev, C. D. Man and a. C. Cobelli, "The UVA/Padova Type 1 Diabetes Simulator Goes From Single Meal to Single Day," *Journal of diabetes science and technology,* vol. 12, no. 2, p. 273–281, 2018.

[101] E. Dassau, J. Pinsker, Y. Kudva, S. Brown, R. Gondhalekar, C. D. Man, S. Patek, M. Schiavon, V. Dadlani, I. Dasanayake, M. Church, R. Carter, W. Bevier, L. Huyett, J. Hughes, S. Anderson, D. Lv, E. Schertz, E. Emory and S. McCrady-Spitzer, "Twelve-

Week 24/7 Ambulatory Artificial Pancreas With Weekly Adaptation of Insulin Delivery Settings: Effect on Hemoglobin A1c and Hypoglycemia.," *Diabetes Care*, vol. 40, no. 12, pp. 1719-1726, 2017.

[102] J. T. Sorenson, *"A Physiological Model of Glucose Metabolism in Man and its Use to Design and Assess Improved Insulin Therapies For Diabetes"~Dissertation*, Massachusetts Institute of Technology, 1985.

[103] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, S. Lukas, V. Thomas and M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiological Measurement*, vol. 25, pp. 905-920, 2004.

[104] C. Paxton, Niculescu-Mizil and S. Saria, "Developing Predictive Models Using Electronic Medical Records: Challenges and Pitfalls," in *AMIA Annual Symposium Proceedings*, 2013.

[105] N. Natarajan, I. S. Dhillon and P. Ravikumar, "Learning with Noisy Labels," in *Proceedings in Neural Information Processing Systems*, 2013.

[106] J. Pearl, Causality: Models, Reasoning and Inference. Second Edition, New York, New York: Cambridge University Press, 2011.

[107] K. Yuan, B. Ying and A. H. Sayed, "On the Influence of momentum Acceleration on Online Learning," *Journal of Machine Learning Research*, vol. 17, pp. 1-66, 2016.

[108] X.-H. Zhou, N. A. Obuchowski and D. K. McClish, Statistical Methods in Diagnostic Medicine, Hoboken, New Jersey: Wiley, 2011.

[109] M. S. Pepe, The Statistical Evaluation of Medical Tests for Classification and Prediciton, New York: Oxford University Press, 2003.

Appendix

Supporting Tables for Chapter 4

Table 4.1

%-of-original 0.7g/kw-BW	A	B	C	D	E
0%	99.965%	0.035%	0%	0%	0%
20%	99.9378%	0.062 %	0%	0%	0%
40%	100%	0%	0%	0%	0%
60%	100%	0%	0%	0%	0%
80%	100%	0%	0%	0%	0%
100%	100%	0%	0%	0%	0%
120%	100%	0%	0%	0%	0%
140%	100%	0%	0%	0%	0%
160%	99.5%	0.213%	0%	0.287%	0%
180%	97.898%	1.147%	0%	0.956%	0%
200%	95.053%	2.658%	0	2.29%	0

Table 4.1 shows the mean percent time in relevant section of the Clark error grid across the 100 UVa/Padova simulator adult subjects under the meal alteration scenarios given in the left column.

Table 4.2

g/kw-BW	A	B	C	D	E
0.14 g/kw-BW	100%	0%	0%	0%	0%
0.28 g/kw-BW	100%	0%	0%	0%	0%
0.42 g/kw-BW	100%	0%	0%	0%	0%
0.56 g/kw-BW	100%	0%	0%	0%	0%
0.7 g/kw-BW	99.558%	0.442%	0%	0%	0%
0.84 g/kw-BW	99.9654	0.035%	0%	0%	0%
0.98 g/kw-BW	97.464%	2.536%	0%	0%	0%
1.12 g/kw-BW	93.801%	6.199%	0%	0%	0%
1.26 g/kw-BW	89.802%	10.198%	0%	0%	0%
1.4 g/kw-BW	85.864%	14.136%	0%	0%	0%

Table 4.2 shows the mean of percent time in the relevant section of the Clark error grid across the 100 UVa/Padova simulator adult subjects with meals totally removed corresponding to percent (given in left column) of the standard 0.7g/kg BW CHO meal at noon scenario.

Table 4.3

%-of-original bolus	A	B	C	D	E
-30%	100%	0%	0%	0%	0%
-20%	100%	0%	0%	0%	0%
-10%	100%	0%	0%	0%	0%

+10%	100%	0%	0%	0%	0%
+20%	99.904 %	0.0956 %	0%	0%	0%
+30%	96.868 %	1.609 %	0%	1.5222 %	0%

Table 4.3 shows the mean percent time in relevant section of the Clark error grid across the 100 Uva/Padova simulator adult subjects under the bolus alteration scenarios given in the left column (boluses given for single 0.7g/kg BW CHO meal at noon).

Table 4.4

%-Basal	A	B	C	D	E
50%	100%	0%	0%	0%	0%
60%	88.487%	11.323%	0%	0.19%	0%
70%	94.185%	5.76%	0%	0.055%	0%
80%	98.507%	1.493%	0%	0%	0%
90%	99.391%	0.609%	0%	0%	0%
110%	99.44%	0.372%	0%	0.188%	0%
120%	90.674%	5.675%	0%	3.652%	0%
130%	74.967%	13.636%	0%	11.397%	0%
140%	66.621%	16.608%	0%	16.771%	0%
150%	66.596%	16.934%	0%	16.469%	0%

Table 4.4 shows the mean percent time in relevant section of the Clark error grid across the 100 Uva/Padova simulator adult subjects under the basal alteration scenarios given in the left column.