LIÈGE université
School of Engineering

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha

Patient-Specific Metabolic Variability and Precision Glycaemic Control in Critical Care

MARCH 2020

VINCENT UYTTENDAELE

A thesis submitted in partial fulfilment of the requirements for the joint degree of

Doctor of Philosophy in Biomedical Engineering,

University of Liège, Liège, Belgium,

University of Canterbury, Christchurch, New Zealand.

Under the supervision of:

Dr. Thomas DESAIVE

Prof. J. Geoffrey CHASE

LIÈGE université
School of Engineering

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha

Patient-Specific Metabolic Variability and Precision Glycaemic Control in Critical Care

MARCH 2020

VINCENT UYTTENDAELE

JURY MEMBERS:

Prof. Pierre DAUBY (Chairman)

Prof. Bernard LAMBERMONT

Prof. Jean-Charles PREISER

Prof. Guillaume DRION

Prof. Thomas DESAIVE (Supervisor)

Prof. Geoffrey CHASE (Supervisor)

The presented research was supported by the FRIA – Fund for Research and Training in Industry and Agriculture (Belgium); the EU H2020 R&I programme (MSCA-RISE-2019 call) under the grant agreement #872488 – DCPM; the NZ National Science Challenge 7, Science for Technology and Innovation; and the MedTech Centre for Research Expertise (CoRE) funded by the New Zealand Tertiary Education Committee.

Abstract

Critically ill patients often experience stress-induced hyperglycaemia. Elevated blood glucose levels are associated with increased morbidity and mortality. Glycaemic control demonstrated improved outcomes for these patients. However, other studies failed to replicate the results, primarily blaming the increased risk of hypoglycaemia and glycaemic variability, both associated with worse outcomes. These confounding outcomes have resulted in acceptance of hyperglycaemia and reduced outcomes, causing ongoing debate on glycaemic control.

The goal of the thesis is to define what makes glycaemic control hard to achieve safely, prove safe, effective control impacts patient outcome, and demonstrate it is possible to achieve safe, effective control for all patients, despite targeting lower glycaemic ranges.

Metabolic variability is the main factor making glycaemic control hard to achieve safely. More specifically, sudden changes in patient-specific response to insulin (intra-patient variability) can lead to severe hyper- and hypo- glycaemia. Novel analysis of model-based insulin sensitivity and its variability clearly showed while inter-patient variability can be significantly different across patients, intra-patient variability is equivalent. Therefore, no patient is harder nor easier to control, and thus all patients should be able to benefit from similar quality of control. In turn, conclusions on glycaemic control from studies failing to do so may be biased due to poor protocol design, rather than physiological factors related to severity and outcome.

Intra-patient variability is still very large, and it is not possible to discriminate more and less variable patients, reducing the quality of control deliverable in practical clinical scenarios. This research developed a novel 3D stochastic model to optimally segregate more and less variable patients based on prior behaviours. This approach enabled significantly improved, and tighter prediction of risks associated with a given insulin and/or nutrition intervention. Clinical trial results in NZ have shown improved control and safety using this new 3D stochastic model.

To demonstrate these outcomes, a clinical trial using STAR, a model-based, patient-specific glycaemic control framework, was designed and implemented at the University Hospital of Liège. Results showed STAR succeeded in providing safe, effective control to virtually all patients, despite targeting lower target bands associated with better outcomes. However, increased workload compared to the standard protocol was identified as a limitation.

Finally, this thesis develops a means to dramatically increase the STAR measurement interval from 1-3 hourly to 1-6 hourly without significantly degrading performance or safety. Virtual trials clearly defined the risk and reward trade-off between control performance, patient safety, workload, and nutrition. This result allows clinical staff to choose from a far wider range of options and approaches to provide safe, effective control, with clearly defined risk trade-offs.

Overall, a series of analyses and clinical trials have shown safe, effective control is necessary to improve outcomes, and can be achieved for all patients. These outcomes are possible using patient-specific, model-based glycaemic control protocols developed in this thesis, which directly account for both intra- and inter- patient variability and reduce workload.

Résumé

Le stress et l'inflammation chez les patients critiques déclenchent une cascade de réactions ayant pour effet une production endogène de glucose anormalement élevée et une résistance accrue à l'insuline, provoquant de l'hyperglycémie. L'insulinothérapie est donc prescrite chez ces patients, dans le but de réduire ces niveaux anormalement élevés de glycémie, associés à des comorbidités multiples. Plusieurs études ont mis en évidence les bénéfices liés au contrôle strict de la glycémie, mais l'augmentation importante des risques d'hypoglycémie et de la variabilité glycémique, tous deux indépendamment associés à des complications sévères, ont ouvert un débat quant aux effets positifs ou néfastes liés à ce contrôle. En effet, bien que des glycémies normales soient davantage bénéfiques pour les patients, des glycémies légèrement plus hautes permettent de minimiser les risques d'hypoglycémie.

Cette thèse tente d'identifier les facteurs impactant la qualité et la sécurité du contrôle de la glycémie, ainsi que de démontrer qu'il est possible d'offrir un contrôle de qualité pour tous les patients.

Un des facteurs principaux rendant le contrôle difficile est la variabilité de la sensibilité à l'insuline. La sensibilité varie d'un individu à l'autre, évolue avec le temps, et est directement responsable des risques potentiels d'hypoglycémie. Dans cette thèse, il est montré qu'alors que la sensibilité à l'insuline entre les patients est différente, la variabilité temporelle est équivalente. Il en résulte que la qualité du contrôle de la glycémie doit être similaire chez tous les patients, et qu'un protocole mal adapté ne permet pas de s'en assurer.

Caractériser la variabilité de la sensibilité à l'insuline est donc primordial dans le contrôle de la glycémie. Dans STAR, un protocole de contrôle de glycémie, cette variabilité est prise en compte grâce à un modèle mathématique, déterminant la sensibilité à l'insuline spécifique du patient, ainsi qu'un modèle stochastique pour en évaluer sa variabilité. Un nouveau modèle stochastique 3D est développé pour améliorer la prédiction de l'évolution de la sensibilité à l'insuline. Il se base sur l'évolution antérieure de de cette variable, et permet de mieux quantifier les risques d'hypoglycémie liés à un traitement spécifique. Un essai clinique en Nouvelle-Zélande a pu montrer une amélioration de la sécurité du contrôle de la glycémie grâce à ce nouveau modèle.

STAR a également été implémenté dans un essai clinique en Belgique, afin de montrer une fois encore qu'un protocole adapté, prenant en compte la variabilité métabolique des patients, peut contrôler la glycémie de manière sécurisée et efficace. Cet essai clinique quantifie également, pour la première fois, l'impact du contrôle de la nutrition, en plus de l'insuline, aux soins intensifs. Les résultats de l'étude montrent une nouvelle fois une augmentation significative de la qualité de contrôle, et un apport nutritionnel beaucoup plus adapté que lorsque la nutrition est laissée à l'appréciation du staff médical.

Enfin, au vu de l'augmentation de la charge de travail induite par des mesures cliniques plus fréquentes demandées par STAR, cette thèse évalue également l'impact lié à l'utilisation d'intervalles de mesure plus longs sur le contrôle de la glycémie. Au travers de simulations, le compromis entre les risques et les bénéfices lié à un suivi moins régulier de la glycémie sont clairement définis, donnant des pistes de réflexion concernant la meilleure stratégie à adopter.

Cette thèse démontre, au travers d'analyses *in-silico* et d'essais cliniques, qu'un contrôle strict, sécurisé, et efficace de la glycémie est non seulement possible, mais indispensable pour tous les patients, puisque la variabilité métabolique est identique et indépendante de leur état.

Acknowledgments

Over the course of this research project, I have been surrounded by numerous people who have made this an incredible and unforgettable experience.

Thank you to my supervisors Thomas Desaive and Geoff Chase, for your support over the past 4 years. I am grateful for the knowledge, experience and passion which you shared with me and for your guidance in moulding me into the researcher I am today. Thank you for your valuable feedback on this project, this joint-PhD would not have been possible without both of you.

Thomas, merci d'avoir cru en moi dès mes années de master et jusqu'à aujourd'hui. Merci pour tes conseils et ta franchise. Tu m'as souvent laissé prendre mes propres décisions tout en les respectant. Au travers de nos discussions, tu as pu me redonner plusieurs fois confiance en moi.

Geoff, thank you for all your care and support for Florence and I during our time in New Zealand. You were always available to answer questions and offer encouragement. I especially enjoyed your interest in discovering Belgian beers. I admire the way you manage your students, lift them up, and always ensure their wellbeing.

Many thanks to Jennifer Knopp, who inspired me to do research. I truly believe you are the model of a high-quality researcher. Thank you for your support, assistance, and all the time you invested in our discussions at the office, conferences, or elsewhere. I am grateful for your guidance in assessing statistics, for your wisdom when I over-stated ideas, and for your help as I developed my English skills.

Thank you to Dr. Bernard Lambermont, Dr. Geoffrey Shaw, and Dr. Nathalie Layos for your support and assistance in setting up clinical trials in Belgium and New Zealand. Merci aussi à Marc Pirotte et Muriel Bayet d'avoir été les personnes porteuses des essais au sein du CHU de Liège. Merci d'avoir veillé si longtemps à inclure des patients, et de m'avoir toujours bien accueilli dans vos unités respectives.

I would also like to thank all the "*minions*" I met in the Centre for Bioengineering in New Zealand. You were all extremely supportive, and it was a great pleasure to spend time with you. Also, special thanks to Farzin Golzar, your friendship is priceless, your sense of humour is epic, you are full of kindness, and you made my stay at UC and in New Zealand unforgettable. Merci aussi à Mona, Sarah, Hatim, Ramin et Simon d'avoir partagé de bons moments ensemble. Merci d'avoir supporté mes visites incessantes dans votre bureau : désolé mais vous êtes juste à côté de la machine à café !

Merci à tous les U, petits et grands, de loin et de moins loin, d'avoir toujours cru en moi. Votre aide, vos conseils, votre bonne humeur m'ont accompagné dans cette épreuve. Maman, Papa, vous avez fait de moi qui je suis et vous êtes une source d'inspiration.

De manière plus générale, merci à toutes les personnes, amis, famille, qui m'entourent et m'ont encouragé lors de ces 4 dernières années. J'ai souvent pu trouver réconfort au travers de nos discussions, de vos conseils, de nos balades, et des bons moments passés avec vous.

Enfin, merci à Florence. Tu m'as toujours soutenu inconditionnellement malgré tous les sacrifices que tu as dû faire. Tu n'as jamais hésité à me faire passer avant tout le reste, toujours en m'encourageant et en me poussant vers l'avant. Si j'en suis arrivé là, c'est aussi grâce à toi. Tu as partagé cette expérience avec moi du début à la fin, avec les hauts les bas. Tu as été mon rayon de soleil lors de toutes nos aventures, et je souhaite que jamais la nuit ne tombe.

Contents

List of Figures

List of Tables

Nomenclature

%ΔSI	Hour-to-hour percentage change in SI
AIR	Acute immune response
APACHE	Acute Physiology and Chronic Health Evaluation
BG	Blood Glucose
CDF	Cumulative Distribution Function
CHO	Carbohydrate
CGM	Continuous Glucose Monitoring
CI	Confidence Interval
CNS	Central Nervous System
EGP	Endogenous Glucose Production
GC	Glycaemic Control
GF	Goal Feed
ICING	Intensive Control Insulin-Nutrition-Glucose
ICU	Intensive Care Unit
IIT	Intensive Insulin Therapy
IQR	Interquartile range
LOS	Length of Stay
NICE-SUGAR	Normoglycemia in Intensive Care Evaluation-Survival Using Glucose Algorithm Regulation
RCT	Randomised Clinical Trial
ROT	Rule of Thumb
SD	Standard Deviation
SI	Insulin Sensitivity
SOFA	Sequential Organ Failure Assessment
SPRINT	Specialised Relative Insulin Nutrition Table
STAR	Stochastic TARgeted
T2DM	Type 2 Diabetes

Chapter 1: Introduction to Hyperglycaemia and Glycaemic Control in the Intensive Care Unit

This chapter introduces the general context of this thesis, presents the major current issues in the field of glycaemic control (GC) in the intensive care unit (ICU), and presents the fundamental problems. It is concluded by an outline of the general organisation of this thesis to address these problems.

1.1. General Context

Metabolism is a complex biochemical process turning nutrition into energy for all the body's functions. Metabolism can be seen as a set of systems, driven by a wide range of chemical reactions, with survival the only aim [1]. One of these systems is the endocrine system, composed of all the organs which secrete hormones. Hormones are chemical messengers secreted in response to specific stimuli. They are used in regulatory signalling and control system, often acting as a (positive or negative) feedback system.

One of the roles of the endocrine system is to control glucose homeostasis. Glucose is the main source of energy for metabolism, and is vital to organs. Some organs, including the central nervous system (CNS), use only glucose as their source of energy [2]. It is thus important for the body to maintain sufficient blood glucose (BG) concentration, or glycaemia, without having too high or too low a level [1]. The metabolism needs to maintain sufficient BG for organs function, while maintaining glucose homeostasis or balance using the glycaemic regulatory system [1, 3].

The glycaemic regulatory system uses two antagonist hormones secreted in the pancreas: insulin and glucagon [1, 3]. While insulin reduces BG levels by promoting catabolic reactions, which break down and store glucose, glucagon increases BG levels by promoting anabolic, glucose raising, reactions [1]. Other hormones, such as glucocorticoids, epinephrine, adrenaline, and growth hormone, also impact glucose homeostasis, and act overall to raise BG levels [4, 5].

Catabolic reactions (glycolysis, glycogenesis, and lipogenesis) are characterised by degradation of glucose to adenosine triphosphate (ATP), for energy production, or transformation of glucose into glycogen, fats, or lipids, decreasing thus metabolic BG levels [1, 3]. These reactions all reduce BG levels. In contrast, anabolic reactions (glycogenolysis and gluconeogenesis) are characterised by the synthesis of glucose, from glycogen or other substrates, often referred to endogenous glucose production (EGP), and act to increase BG levels [1]. These reactions mainly occur in the liver, the muscles, the adipose tissues, and the kidneys.

Normal BG concentrations, or normoglycaemia, range between 4.4-6.1 mmol/L in healthy patients [1]. After a meal, as carbohydrates (CHO) are broken down to glucose by the gut and absorbed in the blood, BG levels increase. Elevated BG levels, called hyperglycaemia, lead to insulin released by the pancreatic cells, to reduce BG levels back to normal ranges. During exercise, the metabolism and muscles require more glucose as they need more energy. Without the effect of glucagon to promote EGP, BG levels would decrease below normoglycaemia. These low BG levels are called hypoglycaemia.

There exists no strict definition or thresholds for hyperglycaemia and hypoglycaemia [6]. However, BG levels between 8.0-10.0 mmol/L are often referred as moderate hyperglycaemia, while BG levels above 10.0 mmol/L (the renal threshold for glucose) are called severe hyperglycaemia. Similarly, BG levels between 2.2-3.9 mmol/L are called moderate hypoglycaemia, and BG levels below 2.2 mmol/L are referred as severe hypoglycaemia. Importantly, long-term hyper- and short-term hypo- glycaemia are independently associated with higher mortality and morbidity in ICU patients, including severe infection, sepsis and septic shock, myocardial infarction, critical-illness polyneuropathy, severe brain damage, and multiple-organ failure, showing the importance of good glucose regulation [5, 7-14].

In everyday life, people with diabetes are example of glycaemic dysregulation [1, 3, 15]. There are two major types of diabetes. Type 1 diabetes is characterised by an insulin secretion failure from the pancreatic cells [1, 15]. Insulin is thus not sufficiently secreted (or not at all). Type 2 diabetes (T2D) is characterised by increased insulin resistance by the cells [1, 15]. Insulin is thus not sufficiently effective to reduce BG levels. While these two pathologies differ, they both lead to sustained hyperglycaemia. Because long-term hyperglycaemia is associated with diverse complications [16-20], these individuals need exogenous insulin to help their pancreas and glycaemic regulatory system maintain glucose homeostasis. However, these individuals are not the only ones suffering from glucose dysregulation.

1.2. Hyperglycaemia in Critically Ill Patients

1.2.1. Stress-induced Hyperglycaemia

During the early phase of critical illness, severe stress can impair the glucose regulatory mechanism in patients with or without known diabetes, caused by a complex interplay of reactions to ensure glucose energy is available to vital organs [3, 5, 8, 13, 21]. The stress response to the insult of critical illness

increases counter-regulatory hormones, such as cortisol, epinephrine and glucagon, catecholamines, and growth hormone, all of which promote anabolic glucose raising reactions [3, 5, 13, 21-24]. In addition, there is overproduction of cytokines, such as tumour necrosis factor alpha (TNF-α) and interleukin-1 (IL-1), both pro-inflammatory mediators, altering the insulin signalling pathway, resulting in higher insulin resistance, and thus, impaired glucose uptake by insulin-mediated cells (muscles and adipose tissues) [3, 5, 21, 22]. Finally, there is a pro-inflammatory acute immune response (AIR), which has similar glucose raising impact [13, 21, 23]. The increased glucose production and insulin resistance in all these cases leads to excessive BG levels [5], called stress-induced hyperglycaemia.

In addition, this hyperglycaemic metabolic state, and thus elevated BG levels, are pro-inflammatory [22], as are stress hormone and AIR responses noted previously. This outcome leads to an important inflammatory response to insult, which can be addressed by the anti-inflammatory effect of insulin lowering BG levels [13]. However, the increased insulin resistance induced by the stress response limits the impact of insulin in reducing this inflammatory behaviour and BG levels. Hence, these behaviours can act as positive feedback loops, leading to further increased glycaemic levels and increased insulin resistance. Thus, these patients can be both significantly hyperglycaemic and equally hyper-insulinemic.

Other contributors, such as pre-existing diabetes, level of nutrition, and medications also play a role [3, 5, 13]. More specifically, glucose regulatory system impairment is generally greater in patients with diabetes as insulin secretion is already impaired and cannot compensate for increased BG [21]. Drugs, such as glucocorticoids, β-blockers or catecholamines, often used in critical care and often diluted with glucose solutions, are known to increase hyperglycaemia and further increase insulin resistance [25, 26]. However, the first two have been shown to have a lesser impact on BG levels in the ICU due to the very high insulin resistance [26]. Finally, inappropriately high nutritional support with excessive CHO content, especially during the early phase of critical illness, can also lead to excessive hyperglycaemia [5, 23, 27-29].

Together, the set of stress and inflammatory responses create self-sustained feedback loops from stress-induced hyperglycaemia (Figure 1.1), which are associated with increased morbidity and mortality [3, 5, 7, 13, 30] and occur in 30-50% of ICU patients [7, 31].

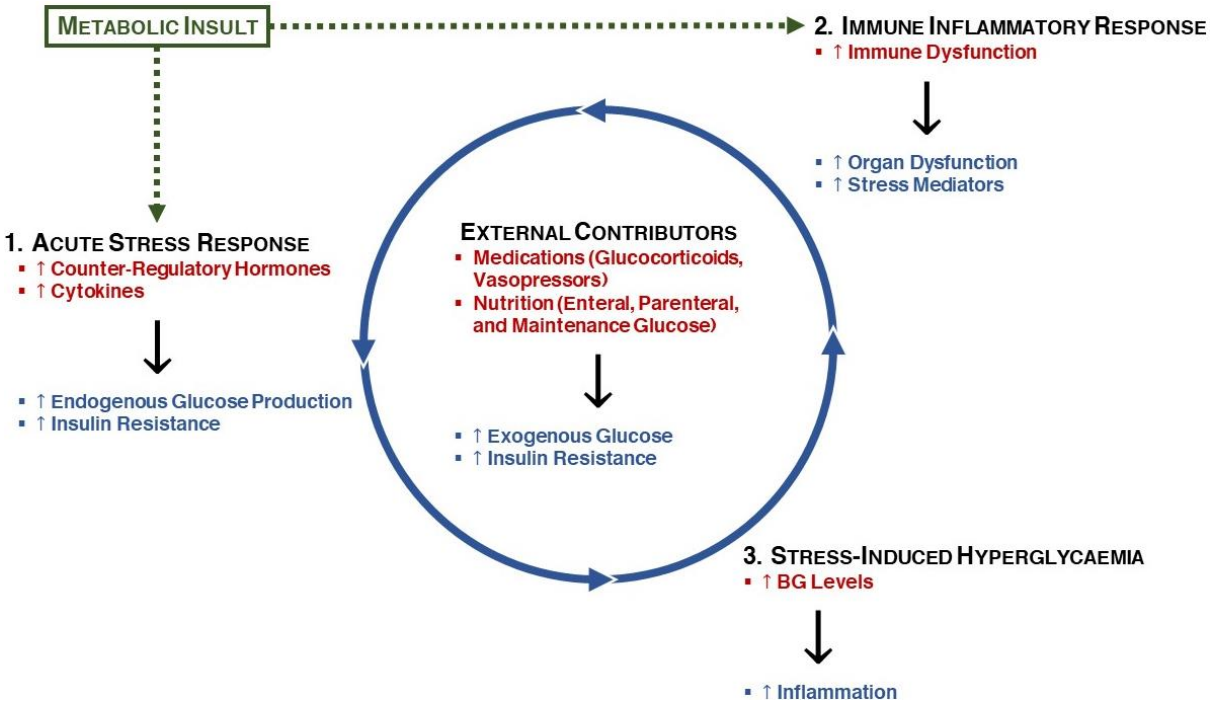

Figure 1.1 – Self-sustained positive feed-back loops of stress-induced hyperglycaemia, where lowering glucose can mitigate these loops.

1.2.2. Breaking the Loop

Due to the association of hyperglycaemia with higher complications [5, 7, 13, 30], it is essential to break this self-sustained or positive feed-back loop driven, stress-induced hyperglycaemia. Given excessive glucose production and high insulin resistance, the only way to break the loop is to use exogenous insulin to lower BG levels. Higher insulin concentrations can account for excessive insulin resistance to lower BG levels, while also providing anti-inflammatory effects. Together, they can break the self-sustained stress-induced hyperglycaemia loops [5, 13].

However, the acute phase of metabolic illness is adaptive and patient-specific [8, 27, 29]. Insulin resistance can thus vary over time, and excessive insulin administration can lead to hypoglycaemia, a critical safety concern [10, 11, 32-34].

1.3. Glycaemic Control in the Intensive Care Unit

1.3.1. Insulin therapy – An Eternal Debate?

GC using insulin therapy is thus used to reduce BG to safer concentrations, and several studies have shown improved outcomes by reducing organ failure, clinical burden, and cost [35-45]. However, other

studies failed to replicate these results, showing increased glycaemic variability and higher risk of hypoglycaemia [46-53], which are both independently associated with severe complications and death [10, 11, 14, 54-59]. The optimal target band for GC has since been strongly debated and remains unclear [60, 61]. More specifically, whether GC protocols should target normoglycaemic ranges (4.4-6.1 mmol/L), or tolerate permissive hyperglycaemia (≥8.0 mmol/L) is the main question [24].

A summary of some clinical studies results and outcomes analysing the impact of using lower or higher target bands on clinical outcome is presented in Table 1.1. Clearly, studies showing improved outcomes and supporting the use of lower glycaemic ranges managed to provide safe (<1.5% of patients experiencing severe hypoglycaemia), and effective (>50% BG in target band) control for nearly all patients. In contrast, other studies, typically opposed to lower ranges, resulted in much higher incidences of hypoglycaemia (>6.9% of patients experiencing severe hypoglycaemia) and lower efficacy (<50% BG in target band and increased glycaemic variability) for more patients. A first level analysis would suggest poor GC safety and efficacy are associated with higher risks, rather than the lower target band itself; specifically, not the band (of control), but the method (of control).

These disparities, and several contradictory published results and methods, led to many editorials and meta-analyses expressing disagreement on the subject [8, 60-69]. To date, the recommendations of moderate, rather than tight, BG targets for GC in ICU [70-74] reflects the “first do no harm” or “uncertainty” principle [60], where the increased risk of hypoglycaemia can be more harmful for the patient than the potential benefits from GC [10, 11, 32]. However, these recommendations are heavily based on studies that failed to provide safe and effective control for all patients [75].

High time in target band, high safety from hypoglycaemia, and low glycaemic variability, all associated with reduced mortality [38, 71, 76-80], reflect controller capacity to provide safe, effective control. They all also indicate this control quality must be consistent over time and most (or all) patients, which only a relatively few studies considering outcome achieved [35, 36, 45, 81]. In contrast, significant inter- and intra- patient metabolic variability makes GC hard to achieve safely and drives glycaemic variability and control safety [82-86].

Table 1.1 – Summary of some existing studies on GC supporting or not supporting the use of lower target bands.

	% patient experiencing hypoglycaemia	%BG in target band	Target Band	Supporting lower glycaemic levels
Van Den Berghe et al. [35]	0.05	/	4.4-6.1	Y
Chase et al. [81]	0.3	54	4.4-6.1	Y
Stewart et al. [87]	1.5	87	4.4-8.0	Y
Mesotten et al. [43]	0.9	67	4.4-6.1	Y
Krinsley et al. [36]	0.34	70	4.4-6.1	Y
Pachler et al. [44]	0.05	/	4.4-6.1	Y
Blaha et al. [88]	0.8	83	4.4-8.3	Y
Finfer et al. [46]	6.9	/	4.5-6.0	N
Preiser et al. [47]	8.7	42.8	4.4-6.1	N
Kalfon et al. [89]	13.2	/	4.4-6.1	N
De la Rosa et al. [48]	8.3	<50	4.4-6.1	N
Brunkhorst et al. [50]	17%	/	4.4-6.1	N

1.3.2. Model-Based Glycaemic Control

Hence, computerised, patient-specific model-based solutions directly accounting for intra- and inter-patient variability are needed to personalise and optimise GC [90, 91]. In contrast, typical table-based "one size fits all" or ad hoc clinical protocols lack patient-specificity and often rely on clinical judgment, introducing variability. Model-based GC methods offer an advantage as they quantify patient-specific metabolism and variability, and consistently modulate insulin and/or nutrition with respect to these time-varying variables [92-97].

More particularly, metabolic models allow identification of patient-specific key physiological parameters reflecting patient metabolic state [91, 98, 99], such as insulin sensitivity (SI) [85], a key parameter describing patient-specific metabolic condition [85, 93, 100-103]. Model-based GC allows greater patient-specificity in treatment and has the advantage of being adaptable across cohorts and clinical practices [104]. They thus offer a "one method fits all" solution providing personalised care. It has been shown to provide tighter, and thus less variable, control, significantly improving clinical outcomes [41, 43, 81, 85, 95, 96, 104-109].

1.4. What Are the Needs?

GC is at a crossroad between permissive hyperglycaemia and the inability to provide safe, and effective control [64]. The negative association of GC with hypoglycaemia and glycaemic variability overbalance the positive association with improved outcome. To benefit patients, GC must be safe, with minimal

incidence of hypoglycaemia, effective, with high time in target band and reduced variability, and replicable across different ICU clinical practices for robustness.

There is thus a critical need to understand and define what makes safe GC difficult, and demonstrate high GC quality is achievable for all patients, despite targeting lower ranges. More specifically, if all patients can benefit equally from safe and effective GC, avoiding thus incidence of hypoglycaemia, clinical outcome associated with the different GC target ranges should be reconsidered. In turn, this would make a case for GC targeting lower BG ranges, and provide evidence for updated guidelines on GC.

This thesis focuses on three main known aspects influencing GC outcome: metabolic variability, compliance to protocol, and the impact of measurement frequency on GC safety. Although these key factors are well known, this thesis aims to better understand, characterise, and assess their true impact and potential implications in GC. More specifically, this thesis aims to answer the four following critical questions:

1. What are the main factors influencing high quality GC, and why did some studies successfully provide safe, effective control, while others did not?

2. Is poor GC due to patient severity and outcome, and thus unavoidable? Or should everyone be able to benefit from equally safe and effective control, regardless of clinical outcome?

3. How can precision GC be achieved for all? And can this patient-specific precision be increased?

4. What is the risk and reward of longer treatment intervals?

Based on the answers to these four questions, this thesis then aims to address the identified issues and provide solutions to improve patient-specific GC for all ICU patients. The analyses and methods presented are developed on retrospective clinical data and validated in pilot clinical trials.

1.5. Preface

Chapter 2 develops the statistical analysis methodology used in this thesis, as well as the metrics used to compare GC outcomes.

Chapter 3 presents and describes the main tools that are used in the context of this thesis. The proven, model-based Stochastic TARgeted (STAR) GC framework is detailed, and the concept of validated virtual trial methods is also developed. Virtual trials enable assessment of simulated virtual patient GC outcomes for different protocols.

Chapter 4 provides an overview of the different retrospective cohorts of patients used for the different analyses.

Chapter 5 aims to provide an answer to the first question, by comparing two protocol designs from contradicting studies. These two protocols both target normoglycaemic ranges, but one managed to provide safe control for nearly all patients, while the other did not.

Chapter 6 provides a first response to the second question, and analyses both inter- and intra- patient variability in survivors and non-survivors. It highlights the importance of patient-specificity and patient variability in protocol design to provide safe GC for all.

Chapter 7 further investigate variability, comparing the differences between males and females. Overall, together with chapter 6, it shows that, while inter-patient variability is never equivalent across patients, intra-patient variability is always equivalent, suggesting all patients should be able to receive equal quality GC.

Chapter 8 develops a first approach to better characterise intra-patient variability and provide more accurate predictions of future metabolic variability. The new methodology presented provides high evidence of potential improved GC, but has some significant resolution limitations.

Chapter 9 provides a more robust methodology to predict variability. In turn, these improved predictions are used in the STAR model-based GC framework to improve personalisation of the risk-based dosing of insulin and nutrition, responding partially to the third question.

Chapter 10 presents clinical trial results of STAR using the new methodology developed in Chapters 8 and 9. Clinical trial results are compared to retrospective data from the original STAR protocol.

Chapter 11 presents clinical trials results comparing STAR in a Belgian ICU, and quantifying for the first time the impact of additionally modulating nutrition, compared to insulin alone, on GC outcome.

Chapter 12 discusses and analyses the impact of measurement intervals on GC outcome and nutrition intake in the context of STAR, to answer the last question. This is done by extending from 1 to 3-hourly to 1 to 6-hourly the measurement intervals used in STAR.

Chapter 13 provides a summary of the key results of this thesis and their implications for GC.

Finally, **Chapter 14** discusses the future steps toward achieving safe, effective GC for all in ICU.

Chapter 2: Data and Statistical Analysis

Statistics are something often considered “scary” and “vague”, as understanding specific meanings and limitations can be difficult and unintuitive. On the other hand, lack of understanding and false confidence in statistical methods can also result in their over-use and miss-application. Hypothesis testing is one area of statistics commonly used in medicine, that is both well used and misused [110-112].

Hypothesis testing to determine whether a difference is statistically significant at a certain significance level α is now extensively used in scientific research. However, there are many erroneous usages and misinterpretations of these statistical outcomes, fed by many misconceptions about data analysis and statistics [111]. A particularly common misinterpretation is $p>\alpha$ means “the same” rather than, or in addition to, “not statistically significantly different”. This chapter thus briefly presents the biostatistical methodology and concepts used in this thesis, which underly much of the later analysis.

2.1. Introduction

In this thesis, many comparisons are made between different groups, proportions, or distributions to answer clinical questions. More broadly, scientific or medical research often uses parametric or non-parametric methods to determine whether a difference between two populations is statistically significant or not, at a certain pre-set significance level α (often α=0.05). However, statistics is a field with many potential pitfalls and perils, stemming from common misconceptions [111], and leading authors and/or readers to erroneous conclusions about the significance of published studies.

One of the most topical and significant issues in the spotlight recently is P-hacking [111, 112]. P-hacking is the well-intentioned by misguided concept of making many comparisons in the hope of finding significant difference in one or two contexts, as shown in Figure 2.1. In broader terms, it is the result of the push for a p-value<0.05, regardless of its true meaning. P-hacking might involve using different statistical tests, analysing (non-motivated) subsets of data, removing outliers, comparing different outcome variables, *etc* [111].

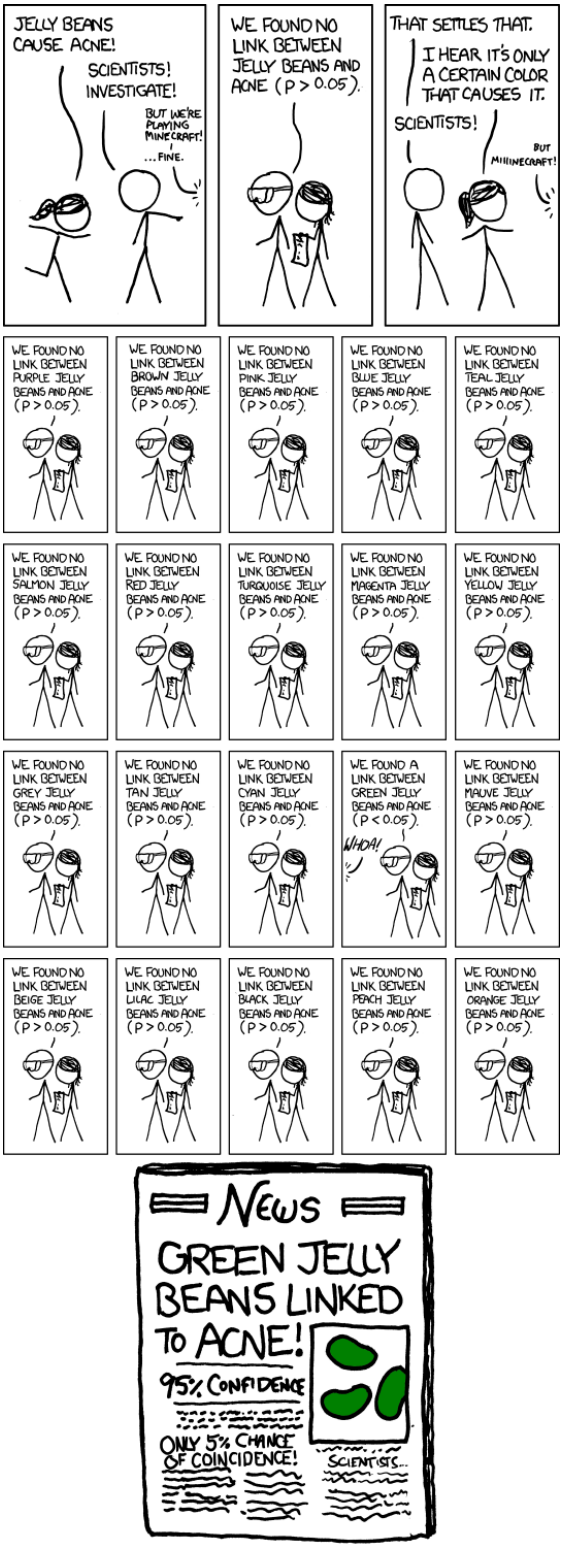

Figure 2.1 - P-hacking demonstrated from https://xkcd.com/882.

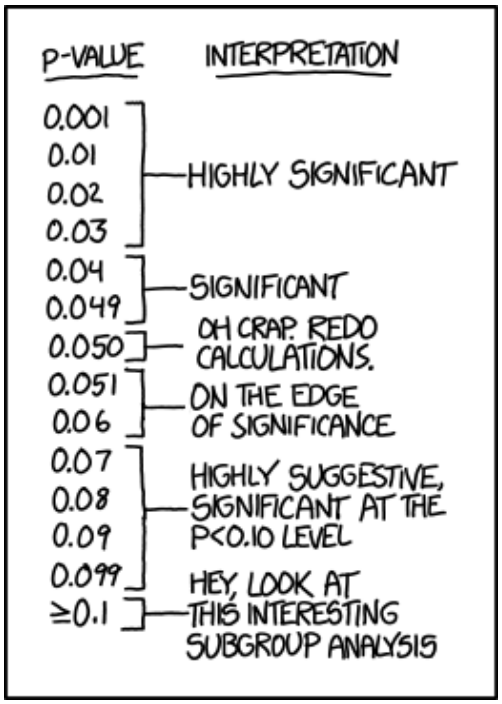

Figure 2.2 – P-values interpretation from https://xkcd.com/1478.

Additionally, the interpretation of p-values is often erroneous, motivated by an obsession around this dichotomous classification (Figure 2.2) [112, 113]. While this value quantifies the confidence of rejecting a null hypothesis (often postulating the absence of difference between independent variables), it does not provide information on the statistical power of the analysis, often forgotten in published papers [110], nor does it necessarily imply that an alternative hypothesis of similarity is true. Overall, it is important to clearly describe and motivate the choice of the statistical methods used and the questions being tested. It gives a clear understanding to what extend the use of such methods are meaningful, but also motivates the conclusions around it.

This chapter thus details briefly statistical methods and analysis, and data representation, used in this thesis. However, this chapter does not aim at demonstrating these concepts. Overall, the choices and methods presented here are heavily based on concepts presented in [110].

2.2. Data Representation

In medical sciences, large clinical datasets are used and compared. In this thesis, patient data may include demographic information (e.g. age, body weight, severity scores), but can also include continuous, time-dependant data, such as the evolution of BG levels over time. With any continuous, or semi-continuous, data set the number of data points can be very large. Hypothesis testing on very large

data sets are more likely to give statistical significance to very small differences between populations, even if that difference is practically or clinically negligible.

2.2.1. Median vs. Mean

When comparing clinical data, the median and interquartile range (IQR) are preferred to the mean and standard deviation (SD), because Median and IQR provides a better intuitive representation, characterisation, and comparison of the overall distributions of two samples, especially where the data is non-normal. The median is also not influenced by large outliers compared to the mean or skewed distributions. In short, it does not assume any distribution shape in the original data. In contrast, reporting the mean and SD metrics are often, but not necessarily, associated with the idea that the data has a Gaussian distribution [110], and, thus, the conclusion that ~96% of the data lies within ±2SD around the mean, which is often not the case with metabolic data. N

For example, a histogram of N=5000 data points sampled from a non-negative lognormal distribution (μ=5e-3, σ=0.7) is shown in Figure 2.3. The mean (SD) of this data sample is $\tilde{\mu} = 1.29 (\tilde{\sigma} = \pm 0.97)$. The SD here implies the sample contains negative values, as $\tilde{\mu} - 2\tilde{\sigma}$ would suggest 2.5% of values are less than -0.65. Thus, if the original distribution is truly unknown, this mean (SD) does not provide any interpretation information about the actual distribution. However, the median [IQR] of this dataset is 1.03 [0.65 1.64], clearly providing information about the “middle” and “central tendency” of this data, which is the goal of this metric. While this issue can be corrected by taking the natural logarithm of the data before calculating the mean, many studies do not or do not state whether they have done so. In addition, a log-normal mean is less intuitive, and again assumes an underlying distribution.

2.2.2. Cumulative Distribution Functions

Cumulative distribution functions (CDFs) are a clear way to compare two samples. It shows the actual distribution of data within the samples, and is the integral of the probability density function which captures the histogram of the data. Thus, CDFs clearly show the median, and all the different percentiles (y-axis) for any given data entry.

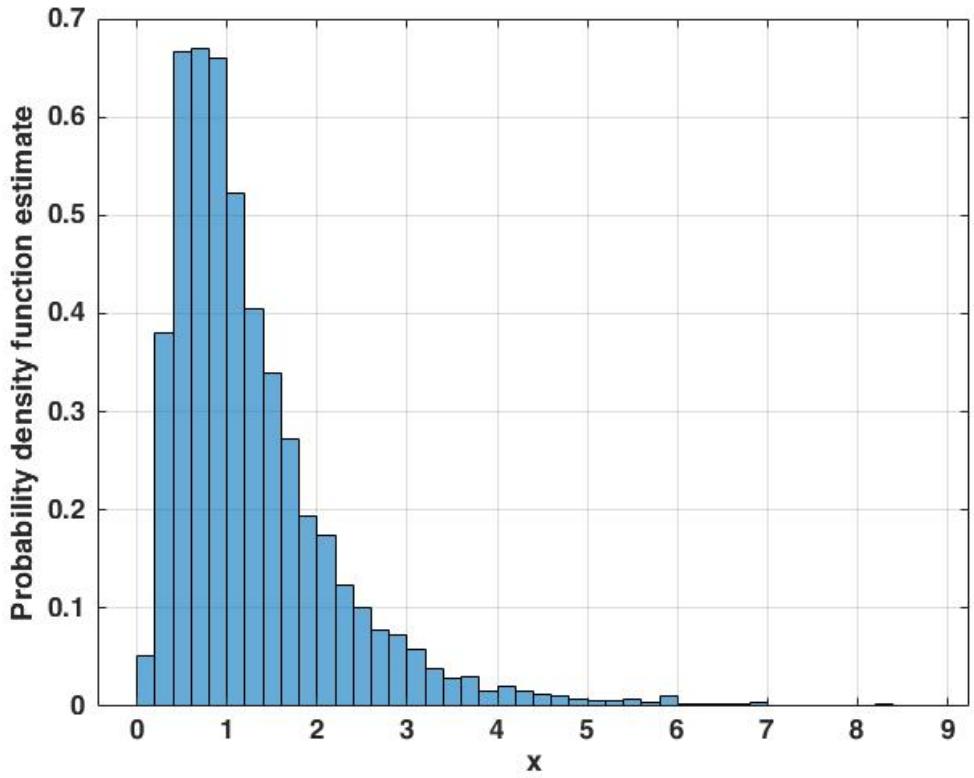

Figure 2.3 – Normalised histogram of N=5000 sampled values from a Log-Normal distribution (μ=5e-3, σ=0.7).

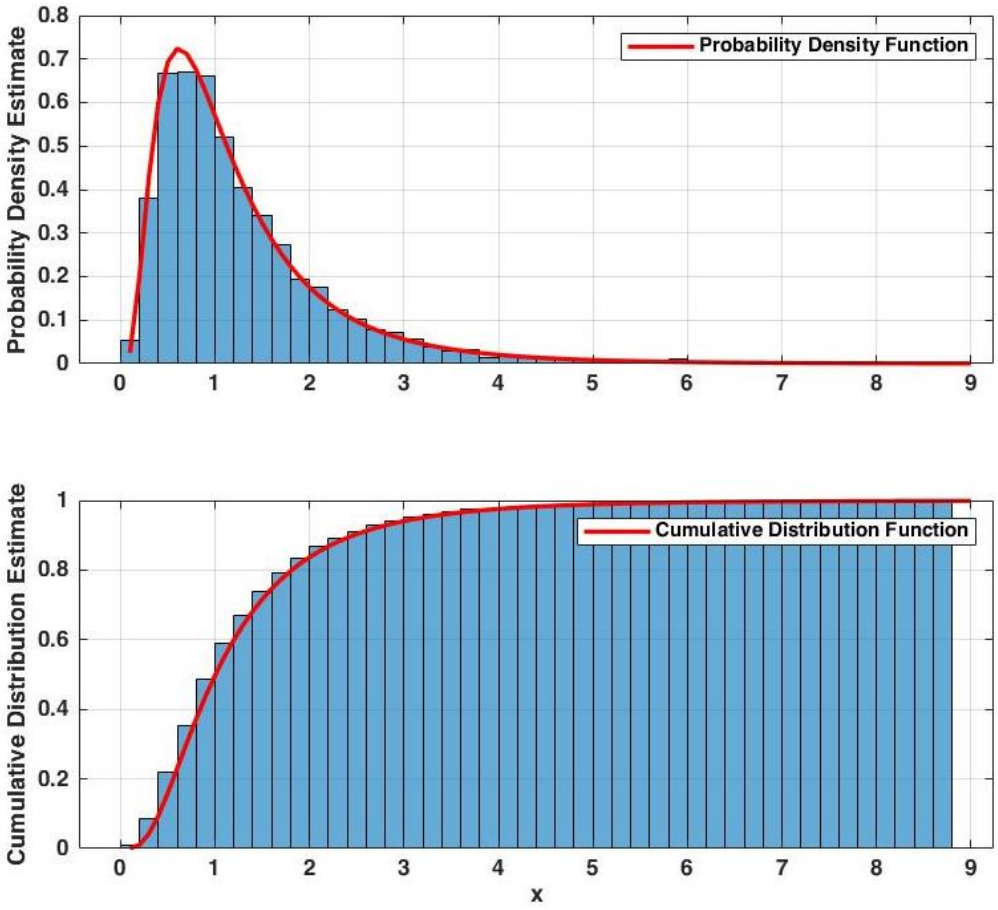

Figure 2.4 – Example of cumulative distribution functions.

From the example presented in Figure 2.3, the corresponding CDF is shown in Figure 2.4 (bottom panel). The median is thus easily identifiable, and located at y=0.5, where is x≃1 (true median = 1.03). Similarly, you can easily estimate any percentile or range of this specific distribution data, and the skewed tail at higher values of x is also evident.

CDFs are easy to create and use as they simply represent the cumulative count of samples for each x-axis values. In addition, they allow very easy comparison of two distribution in terms of median and variability. Figure 2.5 presents different CDDs of samples drawn from different Gaussian distributions, with different mean but equal SD (left panel), and with equal mean but different SD (right panel). Two distributions with different medians have a shift horizontally. Two distributions with different variability (standard error, or variance) will have a different shape.

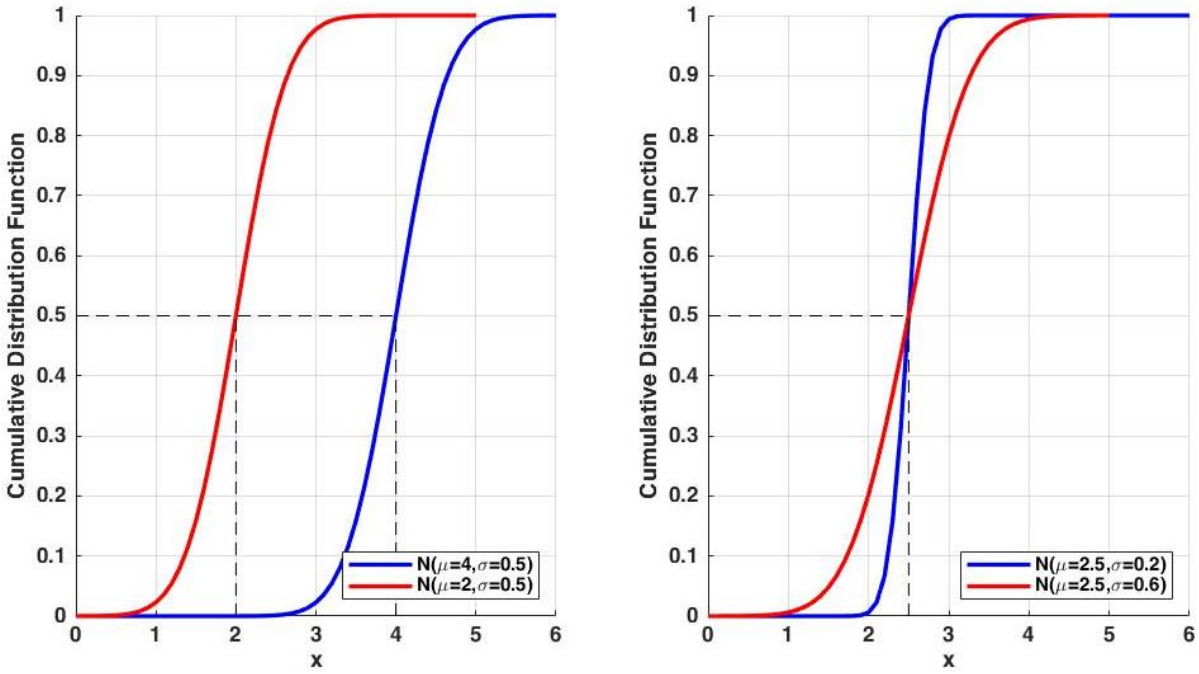

Figure 2.5 - Comparison of CDFs for different Gaussian distributions.

2.3. Hypothesis Testing

In medical research, it is often desired, or even 'required', that a specific change in practice has a statistically significant impact on a specific outcome measure. Thus, the difference between two distributions, often referred as the control and the test distributions, are usually compared using hypothesis testing. In doing so, the aim is to determine whether a null hypothesis (H_0) can be rejected

or not, at a given significance level α. This significance level is often (arbitrarily) chosen as α=0.05, which will be used in this thesis.

There is a wide range of different hypothesis tests, based on the type of data (categorical or continuous data for example), and parametric or non-parametric assumptions. Parametric tests, such as the Student's t-test, are based on assumptions in the distribution of the data, which can thus be defined by parameters, such as the mean and SD. In contrast, non-parametric tests, such as the Mann-Whitney U test or the Fisher Exact tests, are not based on any assumptions on the dataset distribution. In this thesis, non-parametric tests are mainly used, as they do not require any assumption on the data distribution, and they are valid for Gaussian and non-Gaussian distributed data. However, non-parametric methods sometimes suffer from lower statistical power than parametric tests, and may also offer statistically low p-values when N is large [111].

2.3.1. Categorical Data Comparison

In this thesis, categorical (or proportional) data will be compared using the Fisher Exact test, similar to the known χ^2 test, but more robust and valid for any sample sizes [110]. This test analyses the H_0 that the two variables are independent, the alternative being the two variables are not independent. The p-value can be directly calculated using contingency tables.

2.3.2. Continuous Data Comparison

Continuous data is defined as data which can fall anywhere along a spectrum, as opposed to discrete categories. It does not, in this context, represent a 'continuous' data stream in time. Continuous data distributions are compared using two well-known non-parametric statistical tests, based on ranking comparison of the data. The first is the Mann-Whitney U test (also called Wilcoxon rank sum test), analysing whether two independent distributions have equal medians. The second is the Kolmogorov-Smirnov test, analysing if two independent samples are drawn from the same underlying distribution by comparing the distances between points in the sample distributions.

In contrast to the Mann-Whitney U test, the Kolmogorov-Smirnov test can identify both bias in medians, and the shape (variance) difference in distributions. For example, N=100 samples are drawn from each distribution presented in the left panel of Figure 2.5. The two resulting distributions are compared using

these tests. The corresponding p-values using the Mann-Whitney U and Kolmogorov-Smirnov tests are p=0.892 and p=0.002, respectively. Thus, the first test, as expected, does not provide sufficient significance (p>0.05) to reject the H_0 of data being drawn from distribution with equal medians, while the second provides enough statistical significance (p<0.05) to reject the H_0 of the two distributions being drawn from the same underlying distribution (equal medians and equal variance).

This example thus clearly illustrates the importance of choosing, and motivating the statistical test used to show statistical difference. Both of these tests will be used in this thesis to analyse distributions from continuous independent variables. The Mann-Whitney U test will be used if the shape (variance) of the potential distribution is not truly relevant or useful, while the Kolmogorov-Smirnov will be preferred when focusing on variability.

2.3.3. Confidence Intervals

A confidence interval (CI) provides a quantitative way to characterise an unknown general population only based on data sampled from this underlying population [110]. By definition, CIs assume the sample is representative of the population data it comes from. For example, the CI of the median of a population can be estimated from the available observations. One of the main disadvantages of the CI is it relies on many assumptions such as data distribution and independency, and can be very complex to calculate [110].

This CI provides an estimation of how certain it is that the true median of the population lies within in this range, based on the sampled data. For example, a 95% CI provide an interval outside of which there is only 5% chance that the true median of the underlying population lies outside this range [110]. As an example, the resulting 95% CI of samples with different sizes drawn from the Gaussian distribution N(μ=4, σ=0.5) in Figure 2.5 (left panel) are computed and shown in Figure 2.6. The larger the available data, the tighter the CI, since it is based on more information about the original population, and thus there is greater confidence about where true median (or mean) is located.

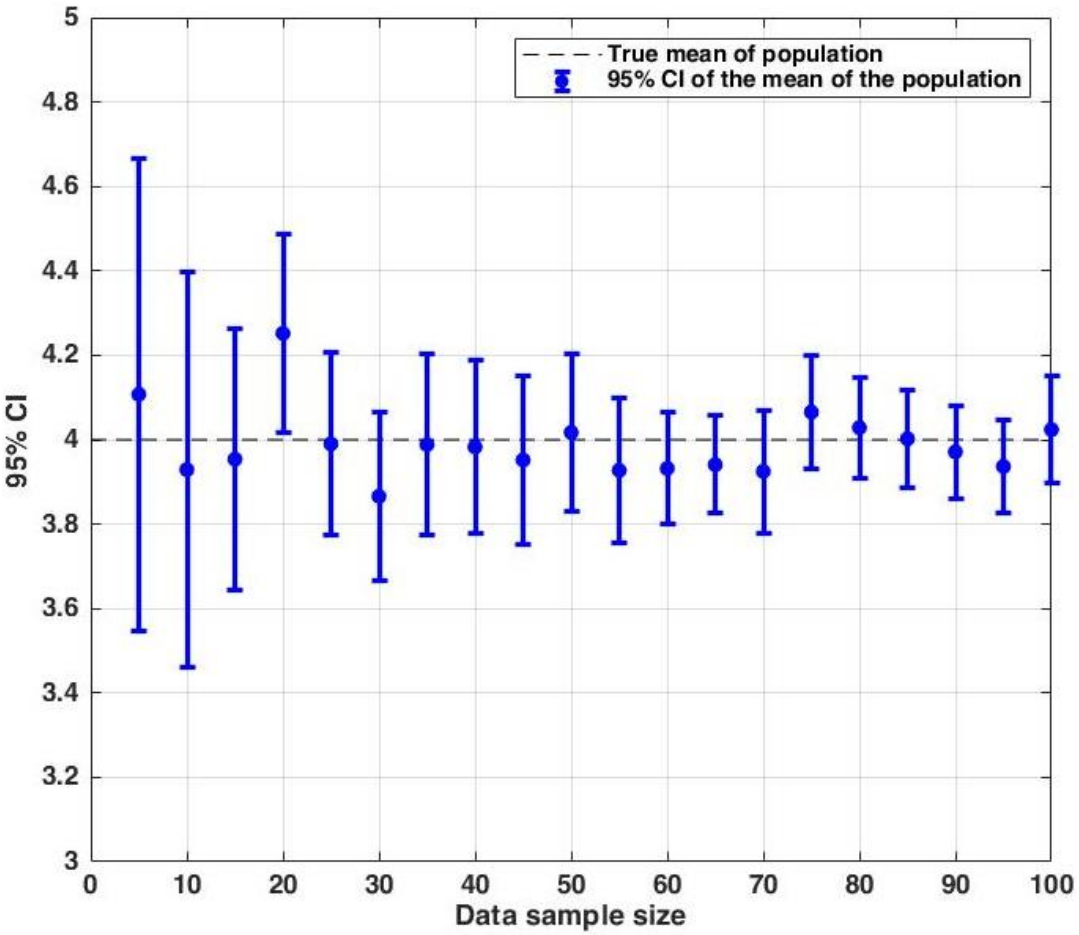

Figure 2.6 – Calculated 95% CI of the mean of the population as a function of data sample size drawn from the Gaussian distribution N(μ=4, σ=0.5).

However, as seen for N=20 in Figure 2.6, by random chance the 95% CI failed to include the true value μ=4. Thus, the 95% CI chosen, similarly to the statistical testing described previously, provides an arbitrarily chosen significance level of α=0.05, and there is thus a 5% chance that the true median is outside this range. As expected based on random chance, in Figure 2.6, it occurred in 1 (5%) in of the 20 cases analysed.

Since the CI can provide a powerful description of a population, based on samples from that population, it can also be used for robust hypothesis testing [110]. For example, considering the H_0 of two distributions having the same medians, and thus the difference in medians being null, the 95% CI of the difference in medians of the two populations can be compared. If the 95% CI of the difference in medians includes H_0, the two distributions are not statistically different ($p>0.05$). However, if the 95% CI does not encompass the H_0, the two distribution can be considered significantly different ($p<0.05$). This method can be applied at any significance level desired.

2.3.4. Large Sample Size Effect

One problem related to hypothesis testing is the p-value dependence on sample size [111]. The easiest way to illustrate this dependence is to consider an example where the Mann-Whitney U test is used to compare samples of different sizes drawn from two close probability density functions. In Figure 2.7, two density functions (left panel) are presented as well as the corresponding CDFs (middle panel). Depending on the context, it is likely that these distributions practically different, especially given the large overlapping shown in the left panel. However, in the right panel, corresponding calculated p-values for different samples sizes clearly decrease as N increases. Typically, regardless of the statistical test used, a small difference in distributions will thus be considered significant for large data sample sizes. As a result, a large p-value is not proof of no effect, and a small p-value is not proof of a large effect [111], when using these tests with larger data sets.

In this thesis, thousands of hours of clinical data are used and compared. Therefore, it can be argued that, given a statistical test, the p-value is very likely to be small even if differences are not clinically significant, influenced by the large data size. A way to avoid this influence is using the bootstrapping method. This approach is also a more explicit and robust means to test a hypothesis, as well, even though it is not commonly used.

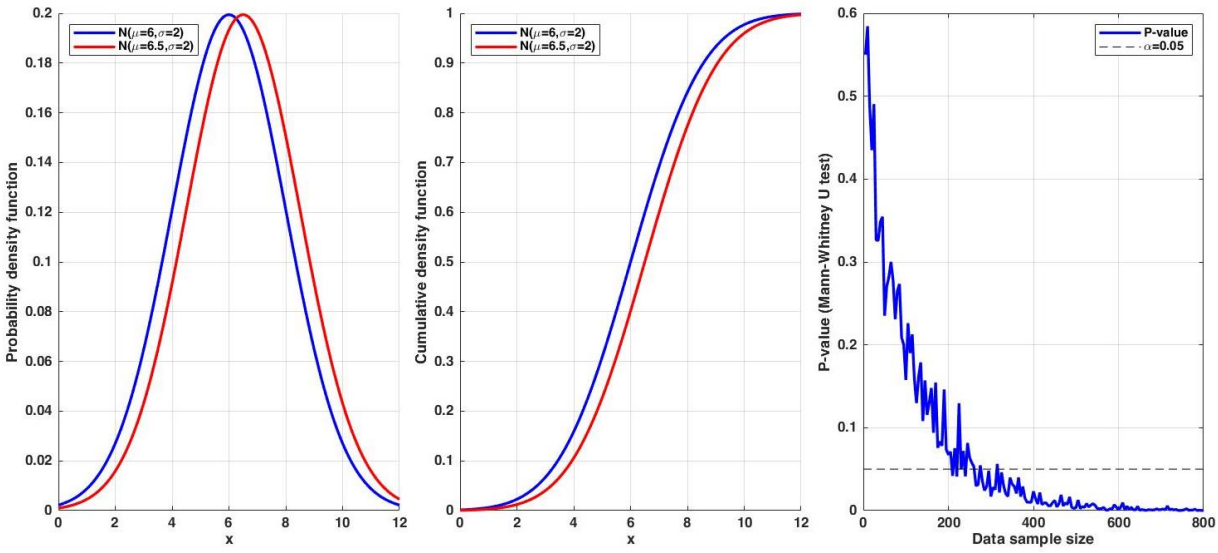

Figure 2.7 – P-value dependency on sample size. P-values are computed using Man-Whitney U test on samples drawn from two close Gaussian distributions.

2.3.5. Bootstrapping

Bootstrapping provides statistic inference of population based on a sample assumed to be representative of that population. Importantly, it does not make any assumption of the original population distribution. The method is depicted in Figure 2.8. The main idea is to calculate the variance of a specific estimate, such as the mean or median of the population, based on resampled samples from available observations. A larger number (N>100) of bootstrap samples are generated by randomly choosing observations, with replacement, from the original sample. Each of the bootstrap samples have the same size as the original sample, and could have the same observation multiple times, or not at all. A bootstrap estimate of each bootstrap sample, such as the mean or the median, can be calculated, providing statistical inference of the true population. Bootstrapping is based on the law of large numbers, where, as $N' \to \infty$, the variance of the bootstrap estimates converges to the true population variance of the studied inference statistic.

The main advantages of this method, in the context of this thesis, is it works for unknown distributions, and avoids the large data size effect from some common (non-parametric) statistical tests [111]. Additionally, it allows to calculate 95% CIs, not requiring any assumptions or complex calculation, which can be used for hypothesis testing (Chapter 2) [114]. This method is also very useful for small available data sets, enabling robust comparison.

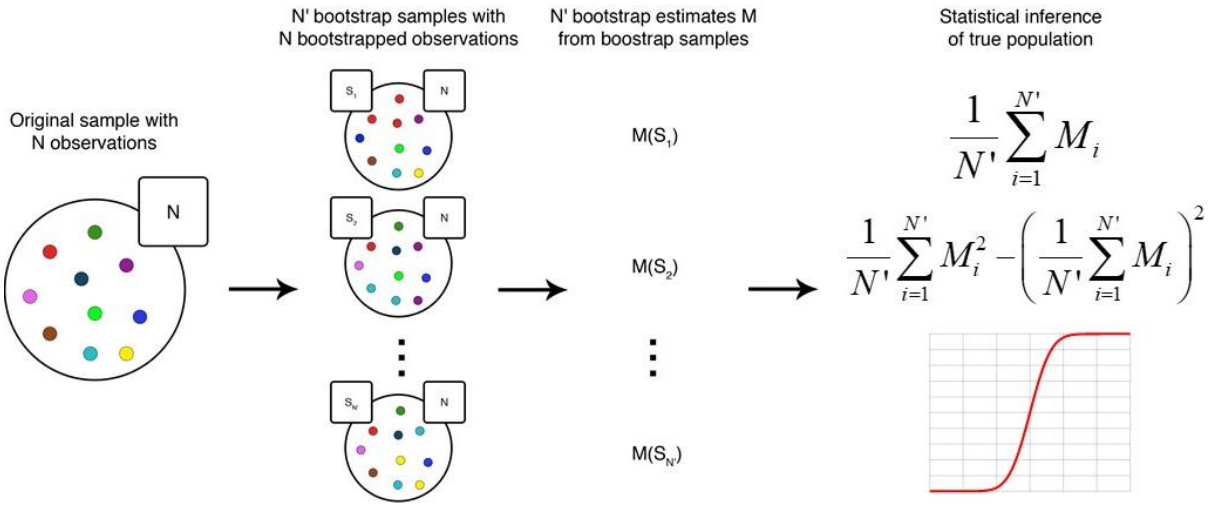

Figure 2.8 – Bootstrapping method summary.

An example is shown in Figure 2.9, where a sample of 10 observations (dark-red dots) were drawn from a true population distribution N(μ=6, σ=2) (solid red), with an estimated mean $\tilde{\mu} = 6.46$ (dashed black). Bootstrapping results provide means (light blue dots) from 100 bootstrapped samples of size N'=10, as the original data sample size. The corresponding 95% CI of the population mean (light blue error bar) is simply calculated by taking the 2.5th and 97.5th percentile of the bootstrapped means. There is 95% confidence that the true population mean (μ=6) lie in this range, which is the case here.

Equally, to compare two populations median values, each pair of sub-sample medians can be compared by subtraction. If the 95% CI of bootstrap estimate difference is < 0, the first set median is significantly lower than the second ($p<0.05$). In contrast, if the 95% CI is > 0, the first set median is significantly higher than the second ($p<0.05$). Finally, if the 95% CI encompasses 0, there is no statistically significant difference in the medians of each set ($p>0.05$). As noted, any significance level α can be tested.

An example considering N=10 observations from each distribution presented in Figure 2.7 is presented in Figure 2.10. The H_0 of distributions having equal medians is considered for statistical testing. The 95% CI of difference in means can thus computed to determine whether H_0 can be rejected at a significance level α=0.05. Since the 95% CI of difference in means include H_0, the null hypothesis cannot be rejected ($p>0.05$).

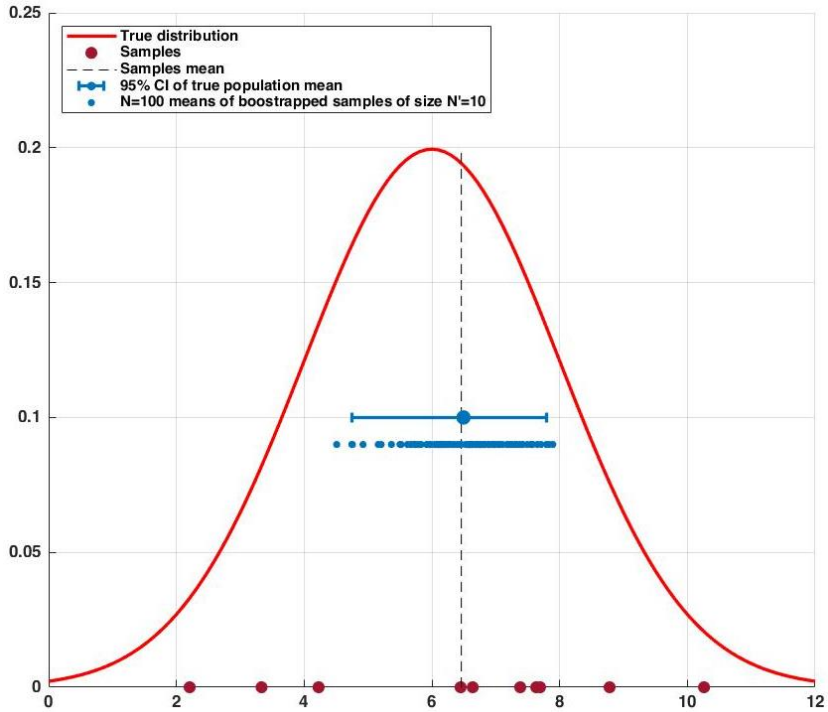

Figure 2.9 – Calculation of the 95% CI using bootstrapping methods.

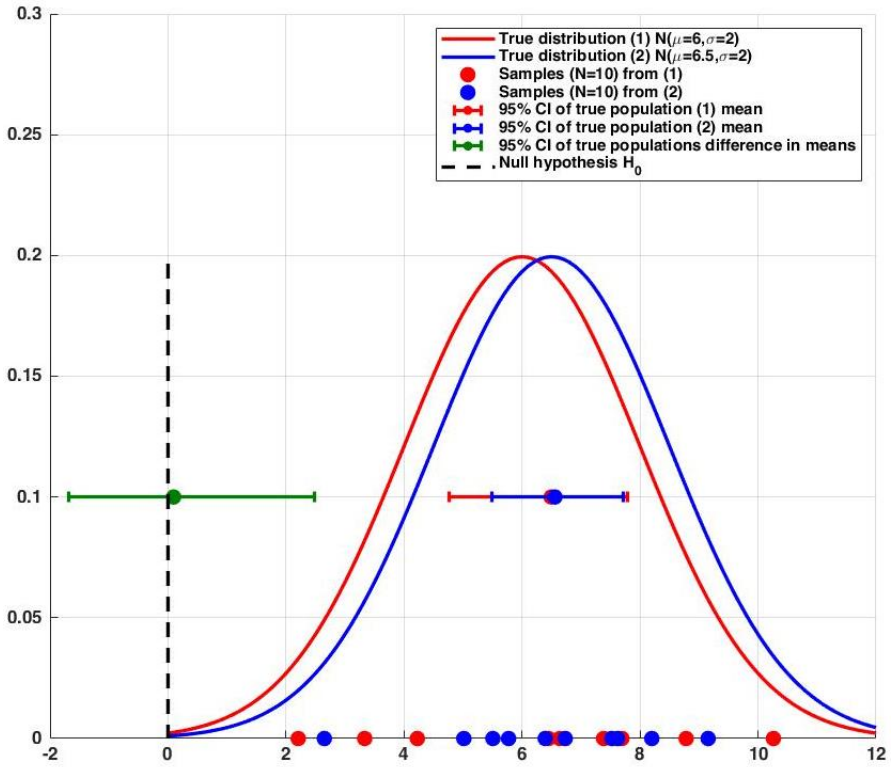

Figure 2.10 – Hypothesis testing using bootstrapped 95% CI of difference in means.

2.3.6. Multiple Comparisons Problem and Bonferroni Correction

The problem with multiple comparisons is that a statistically significant result will be more likely simply by random chance, due to the higher number of comparisons [110]. Thus, it is more likely to make a research conclusion of significance, when the outcome was simply a function of random chance. More specifically, if a statistical test is significant to a level α=0.05, the H_0 can be rejected, with maximum probability of 5% that this rejection was incorrect if H_0 was true (Type I error). If two comparisons on the same data are realised and result in statistical significance, the probability that at least one of the p-values<0.05 by chance is $1.0-0.95^2$= ~10%. Thus, for larger multiple comparisons, the probability to observe one p-value<0.05 significantly increases (Figure 2.1). This problem rises from multiple comparisons of the same family of comparisons on (subgroups of) the same dataset, where comparisons are not independent. The overall concept also applies to CIs.

Thus, to account for these multiple comparisons, the significance level applied to the family of comparisons is adapted using the Bonferroni correction [110]. This conservative method removes the assumptions of comparisons independence. Overall, to ensure there is only 5% (α=0.05) to obtain one or more significant tests, each individual comparison is considered significant to an adapted p-value

equal to $\frac{\alpha}{n}$, n being the number of comparisons. Thus, if 12 multiple comparisons are considered, each individual comparison is considered significant if p < $\frac{0.05}{12}$ = 0.0042.

2.4. Equivalence Testing

Hypothesis testing is used to examine difference between cohorts, and assemble evidence to reject the null hypothesis of data being drawn from the same underlying distribution (or any other H_0). However, it cannot provide evidence for equivalence, especially for large sample sizes [111, 112]. Equivalence testing is used to assess the impact of these differences on clinical decision making, irrespective of the underlying statistical significance (p-value) [110, 115]. Equivalence testing somewhat reflects the power of a statistical difference. However, the equivalence range must be very precisely scientifically determined and motivated.

Thus, it is important to note a difference can be statistically significantly different, but also equivalent, as the first is a statistical measure and the second is a measure of the clinical impact of the difference in the two distributions. Figure 2.11 presents a typical example showing equivalence and statistical significance between two distributions represented by their medians and 95% CI.

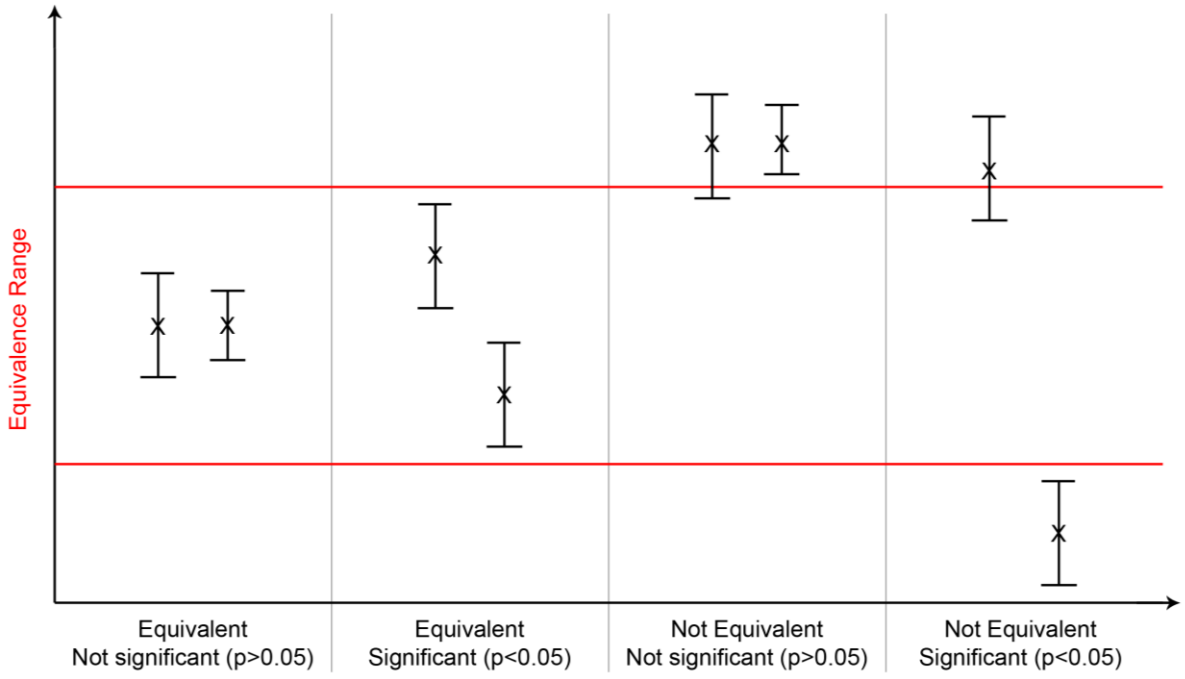

Figure 2.11 – Statistical significance and clinical equivalence between two distributions. Two distributions can be statistically significantly different, but clinically equivalent.

2.5. Reporting Metrics/Statistics in Glycaemic Control

Many statistics and metrics exist to assess GC performance, and all have advantages and disadvantages or limitations [99, 116-119], but have been shown insufficient and inconsistent [117]. This lack of standardisation creates disparities, complicating interpretation and comparison of clinical trials [99, 118, 120, 121].

Table 2.1 presents the metrics mainly used in this thesis to assess and compare GC safety, performance, and workload. These metrics are based on most commonly used, and representative metrics found in the literature [99, 118]. It is important to note the list presented in Table 2.1 is not exhaustive, and other metrics may be used for easier comparison with other published studies or to show other specific outcomes where relevant. Both population and per-patient metrics are used as they are complementary. Per-patient metrics data is always reported as the median [IQR] of per-patient median. Importantly, BG is resampled hourly to allow fair comparison between protocols [87, 122].

Table 2.1 – Metrics used to compare GC safety, performance, and workload outcomes.

Comparison	Metrics	Comments
General	Demographic data such as Number of patients, Age, Sex, Severity of injury, *etc.*	
Workload	Hours of GC	
	Total BG measurements	
	Workload (measurements per day)	
Performance	BG levels	Normoglycaemic levels are associated with improved outcomes.
	%BG in target band (4.4-8.0 mmol/L)	Assesses performance but also reflects protocol design efficacy.
	%BG in 4.4-6.5 mmol/L	High % time in lower BG levels are associated with improved outcomes.
	%BG in 4.4-7.0 mmol/L	
Safety	%BG in 8.0-10.0 mmol/L	Moderate and severe hyperglycaemia are associated with morbidity and mortality.
	%BG > 10.0 mmol/L	
	%BG < 4.4 mmol/L	
	%BG < 4.0 mmol/L	Moderate et severe hypoglycaemia are associated with severe complications and death.
	%BG < 2.2 mmol/L	
	Number of patients < 2.2 mmol/L	
Others	Insulin rates (U/h)	
	Nutrition rates (g/h)	%GF is preferred as it normalises comparison between patients.
	Nutrition rates (%GF)	

2.6. Summary

While p-values are being extensively used in scientific research, the lack of detailed methodology and motivations around it sometimes make the significance of the outcome results, and the conclusions questionable. Large p-value is not proof of no effect, and a small p-value is not proof of a large effect. It is thus important to provide a clear interpretation of outcome p-values, as well as the context in which this value was calculated.

The statistical concepts briefly presented in this chapter will be used in this thesis. More specifically, data will be preferentially reported as median [IQR], and presented in CDFs where appropriate. Hypothesis testing will use either the Fisher exact test for categorical data, the non-parametric Mann-Whitney U test to compare medians of distributions, the non-parametric Kolmogorov-Smirnov test to compare variability (as it compares distributions functions as a whole, not only medians or means), or using the 95% CI of the distribution calculated on bootstrapped samples. The significance level considered will always be $\alpha=0.05$, unless a Bonferroni correction is necessary for multiple comparisons. Finally, the impact of a difference, regardless of its statistically significance, will be assessed using equivalence testing when appropriate.

Chapter 3: STAR – A Personalised, Model-Based, and Risk-Based Dosing Approach for Glycaemic Control

This chapter develops and details the Stochastic Targeted (STAR) GC framework and the use of virtual trials, tools that will be extensively used in this work. The STAR GC framework is a proven model-based and risk-based insulin and nutrition dosing approach accounting for both inter- and intra- patient variability [87, 95, 96, 123]. This GC framework protocol uses a mathematical model to identify patient-specific metabolic state [93], and a stochastic model [124-127] to assess its potential variability in the next 1-3 hour treatment intervals. The risks associated with a given treatment can thus be assessed, and treatment adapted to optimise BG outcomes. Virtual trials are used to simulate GC protocols *in-silico*. Such trials provide accurate prediction of protocol results, enabling comparison and the *in-silico* evolution of safety and performance for different GC protocols.

3.1. Introduction

Chapter 1 presented the importance of personalising GC in critically ill patients. More specifically, there exists no universally accepted clinical best-practice for GC, and most ICUs titrate insulin based on ad-hoc or fixed protocols lacking patient-specificity and relying heavily on clinical judgement [98]. Equally, there are no accepted consensus clinical goals or approaches to GC.

Model-based GC methods offer unique advantages in directly quantifying patient metabolic status and variability, to modulate insulin and/or nutrition [93-96, 123]. Accurate metabolic models allow identification of patient-specific physiological parameters from real-time measurements, such as SI, a key parameter describing patient-specific metabolic condition [85, 93, 103]. Model-based GC thus allows greater patient specificity in treatment and can provide much greater consistency across patients, cohorts, and clinical practices [104]. Thus, the few model-based GC protocols available are able to provide tighter, less variable control than other methods, significantly improving clinical outcomes [41, 43, 81, 85, 88, 95, 96, 104, 107, 109, 123].

Another main advantage of such personalised models is the possibility to design, develop, adapt, and validate GC algorithms *in-silico* [91, 99]. GC protocol design can thus be optimised using virtual simulations, avoiding confounding results of trial-and-error clinical protocol designs [99]. However, this *in-silico* design can only be undertaken providing validated, robust, and generalisable virtual trial schemes using virtual patients, which accurately reflect patient-specific metabolic status [91, 99, 128]. Few models provide validated capabilities [91].

The clinically proven STAR GC framework [87, 95] is an example of a successful, model-based, patient-specific approach, targeting lower glycaemic ranges (Chapter 3). STAR uses a physiological model and a stochastic model to provide a tailored risk-based dosing approach of insulin and nutrition. STAR is the successor of the Specialised Relative Insulin Nutrition Table (SPRINT) [38, 81], originally developed in New Zealand, and has been designed using validated virtual trials[106, 129]. STAR is now the standard of care at the Christchurch Hospital of New Zealand [95, 96, 123], and the Kalman Pandy Hospital ICU, Gyula, Hungary [87].

3.2. STAR Framework

3.2.1. A Unique Approach

Overall, STAR is a model-based, patient-specific GC framework [95, 123]. STAR uses a clinically validated physiological model [93] along with a stochastic model [124, 125] to provide risk-based dosing approach [95, 123]. Inter-patient variability is assessed by identifying model-based, patient-specific SI from patient data [130, 131]. The STAR stochastic model then predicts a distribution of likely future SI for 1-3 hourly intervals, directly quantifying the intra-patient variability of future SI evolution. STAR can thus adjust insulin and nutrition treatment choices to enable a pre-set, clinical risk of 5% of future corresponding BG below any clinically pre-set target band lower limit.

STAR is thus unique in providing a risk-based dosing approach and modulating both insulin and nutrition, where no other known GC protocol, or other drug delivery protocol, does so. Nutrition is reduced if insulin alone is not sufficient to reduce excessive BG levels. Typically, highly resistant patients quickly reach insulin saturation effects on BG uptake [132, 133]. For those patients, nutrition must be reduced to lower BG to safe levels. Thus, nutrition can be temporarily reduced if insulin alone is not sufficient to safely control BG into the target band. STAR uses this capability for safe GC, as well as to maximise nutrition delivery to world leading levels [87, 134].

3.2.2. Insulin Sensitivity

SI is a key parameter describing patient-specific metabolic condition. It characterises the effect of insulin on the glucose regulatory system, reflecting the ability of cells to uptake glucose. SI is the inverse of the more commonly used term insulin resistance.

Higher SI indicates less insulin is required to lower BG levels due to higher insulin-mediated glucose uptake by the cells and tissues. In contrast, lower SI suggests increased insulin is required to provide similar effect on BG levels, since glucose uptake is inhibited by insulin resistance. Equally, higher SI allows more nutrition, and lower SI may require some nutrition restriction.

As introduced in Chapter 1, critically ill patients suffer from increased insulin resistance from stress and inflammatory metabolic response to injury [5]. Over the first ~7 days after ICU admission, the metabolism

experiences different phases of critical illness [27]. A more acute phase, where insulin resistance is at the highest, followed by a recovery phase where insulin resistance progressively decreases. Thus, SI is very low after ICU admission and progressively increases back to normal after some time [82, 135]. SI thus varies over time and across patients depending on the severity of injury and rapidity of recovery, and has been shown treatment independent in the model used by STAR [94, 128].

STAR's risk-based dosing approach uses this key parameter to provide safe control, as it directly influences GC outcomes [95, 96, 123]. Patient-specific SI is identified from a validated physiological model of the glucose-insulin pharmacokinetics [93]. Its validity in the context of this thesis is further explained later in Chapter 6.

3.2.3. Physiological Model

The physiological model used in STAR is the validated Intensive Control Insulin-Nutrion-Glucose (ICING) model [93]. This mathematical model describes the glucose-insulin pharmacokinetics, and has been extensively validated in SI testing and similar clinical studies [130, 136-139]. The glucose-insulin dynamics are represented by a three compartment model, accounting for the appearance of insulin and glucose in blood and interstitial fluid volumes. This compartment model is schematically represented in Figure 3.1, where each metabolic pathway characterises a specific glucose or insulin appearance or clearance. Appearance and clearance rates are represented by incoming and outgoing arrows, respectively.

The glucose compartment (G) is influenced by insulin-mediated and non-insulin mediated glucose transfers. Insulin-mediated clearance transfer is determined by SI and occurs in the liver and in adipose and muscle tissues. Non-insulin mediated transfers include the absorption through exogenous inputs (enteral and parenteral nutrition), the EGP, the kidney clearance, and, the CNS uptake.

There are two compartments modelling insulin exchange and transport. The first is plasma insulin (I) and the second is interstitial insulin (Q). Plasma insulin appearance can come from exogenous insulin, or endogenously secreted by the pancreas. Plasma insulin is cleared through the kidneys and the liver. There is a transportation from plasma to interstitial fluid, where insulin actively impacts insulin-mediated glucose uptake, or is degraded.

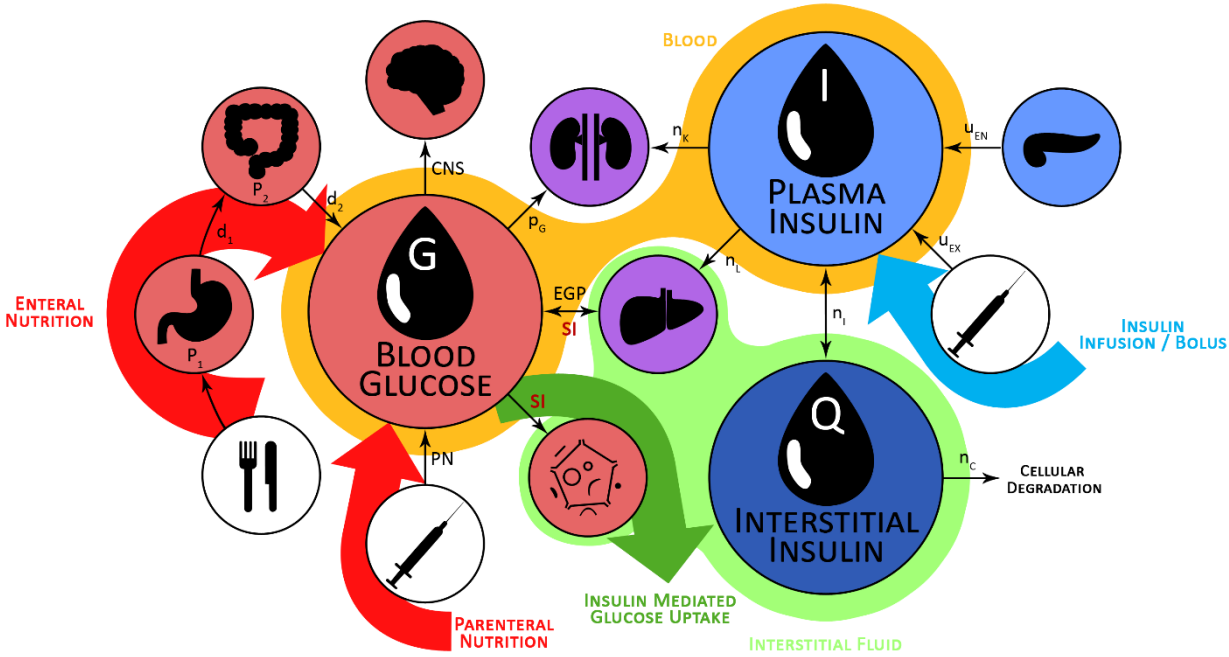

Figure 3.1 – Schematic representation of the glucose-insulin model, showing the physiological compartments and clearances, as well as the appearance of exogenous insulin and carbohydrate, and their kinetic pathways.

Table 3.1 – Key variables of the Intensive Control Insulin-Nutrition-Glucose (ICING) metabolic glucose model (full table in Appendix I).

Parameter	Value	Description
p_G	0.006 min^{-1}	Non-insulin mediated glucose clearance
α_G	1/65 (0.015) l/mU	Saturation of insulin-mediated glucose uptake
EGP	1.16 mmol/min	Endogenous Glucose Production (hepatic)
CNS	0.3 mmol/min	Glucose uptake by Central Nervous System
V_G	13.3 L	Glucose distribution volume
n_K	0.0542 min^{-1}	Kidney clearance of insulin
n_L	0.1578 min^{-1}	Liver clearance of insulin
α_I	1.7x10^{-3} l/mU	Saturation of hepatic insulin clearance
n_I	0.006 min^{-1}	Insulin diffusion between plasma and interstitium
n_C	0.006 min^{-1}	Cellular degradation of internalised insulin
x_L	0.67	Fractional first pass hepatic insulin clearance from portal vein
V_I	4.0 L	Insulin distribution volume
$u_{ex}(t)$	mU/min	Exogenous insulin
$u_{en}(G)$	[16.7 - 266.7] mU/min	Endogenous insulin

The set of three ordinary differential equations models these insulin-glucose dynamics as follow:

$$\dot{G} = -p_G.G(t) - S_I.G(t)\frac{Q(t)}{1+\alpha_G.Q(t)} + \frac{P(t)+EGP-CNS}{V_G} \quad (3.1)$$

$$\dot{I} = -n_K.I(t) - n_L\frac{I(t)}{1+\alpha_I.I(t)} - n_I\big(I(t)-Q(t)\big) + \frac{u_{ex}(t)}{V_I} + (1-x_L)\frac{u_{en}(G)}{V_I} \quad (3.2)$$

$$\dot{Q} = n_I\big(I(t)-Q(t)\big) - n_C\frac{Q(t)}{1+\alpha_G Q(t)} \quad (3.3)$$

where G(t) is the BG concentration (mmol/L), I(t) and Q(t) are the plasma and interstitial insulin concentrations (mU/L), P(t) is the glucose appearance in plasma from enteral and parenteral dextrose intakes (mmol/min), and SI is the insulin sensitivity (L/mU/min). Other parameters are listed in Table 3.1.and the full model details and physiological relevance are presented in the Appendix I.

SI units (L/mU/min) are consistent with a rate parameter assessing the rate of insulin mediated glucose removal, where L/min is a rate of flow or uptake, and the mU^{-1} makes it dependent on current insulin concentration. When SI is multiplied by the average hourly glucose for the period over which SI is calculated, the units becomes consistent with those used in the gold-standard hyperinsulinemic, euglycaemic clamp assessment of SI [138, 140].

The only unknown parameter in Equations (3.1)-(3.3) is SI. All other parameters are known clinically or estimated from population data [97, 125, 141]. Therefore, SI can be identified hourly from clinical BG, insulin, and nutritional data using integral based fitting methods [130, 131].

3.2.4. Stochastic Model

To account for intra-patient metabolic variability, and thus assess unexpected potential changes in metabolic response to insulin, a stochastic model predicting likely future 1-3 hourly changes in SI level (SI_{n+1}, SI_{n+2}, SI_{n+3}) was introduced [124, 125]. These predictions are based on current identified patient-specific metabolic condition (SI_n). This stochastic model was built using a bi-variate kernel density estimation method on population data. The kernel density estimation method enables high resolution behaviour estimation of a specific parameter based upon its prior evolution or state, even where specific data points may be scarce [142]. Details on the development of this methodology and how to provide conditional probabilities based on local data densities are presented in Chapter 9.

For each SI_n value, the stochastic model thus provides a probability density function of likely future SI evolution, reflective of intra-patient variability. In the context of STAR, the 5^{th} and 95^{th} percentiles of these distributions are used to assess the risks associated with a given treatment [95, 125]. More specifically, greater metabolic variability translates to greater potential outcome glycaemic variability in response to insulin, and, thus, potential greater risk of hypoglycaemia. An example of the stochastic model is depicted in Figure 3.2.

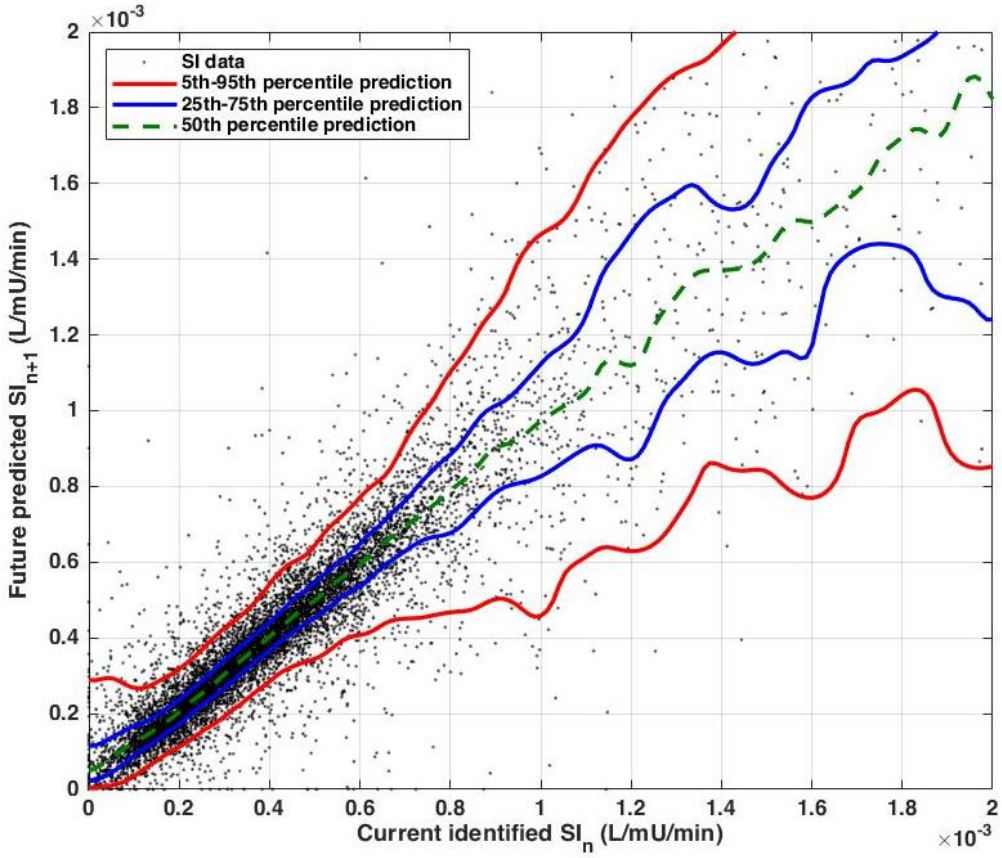

Figure 3.2 – Risk-based dosing approach of the STAR framework. Current patient-specific identified SI is used to forecast the likely 5th-95th percentile range of future SI. This range is used to calculate the corresponding 5th-95th percentile range of likely future BG outcome for a given insulin and nutrition inputs.

3.2.5. STAR Risk-Based Dosing Approach

The STAR treatment selection process uses 4 main steps to determine the optimal, personalised, combination of insulin and nutrition inputs to administer to patients. For STAR, it is set so the risk of BG < 4.4 mmol/L is limited to a maximum of 5%. These steps are graphically represented in Figure 3.3, and further detailed below.

3.2.5.1. Step 1: Identify SI from clinical data

The first step is to identify patient-specific SI_n. The ICING model and integral-based methods are used using clinical data (BG measurement, exogenous insulin rates, and nutrition rates) to evaluate patient current SI level [93, 130]. This step accounts for inter-patient variability, as SI will be unique across patients and hours, based on specific patient response to treatment and patient-specific condition. Importantly, SI is considered constant hourly and treatment independent. For example, in Figure 3.4, $SI_{n=2}$ is calculated based on BG measurements (top panel), insulin (second panel), and nutrition (third panel) rates administered between hours 1 and 2.

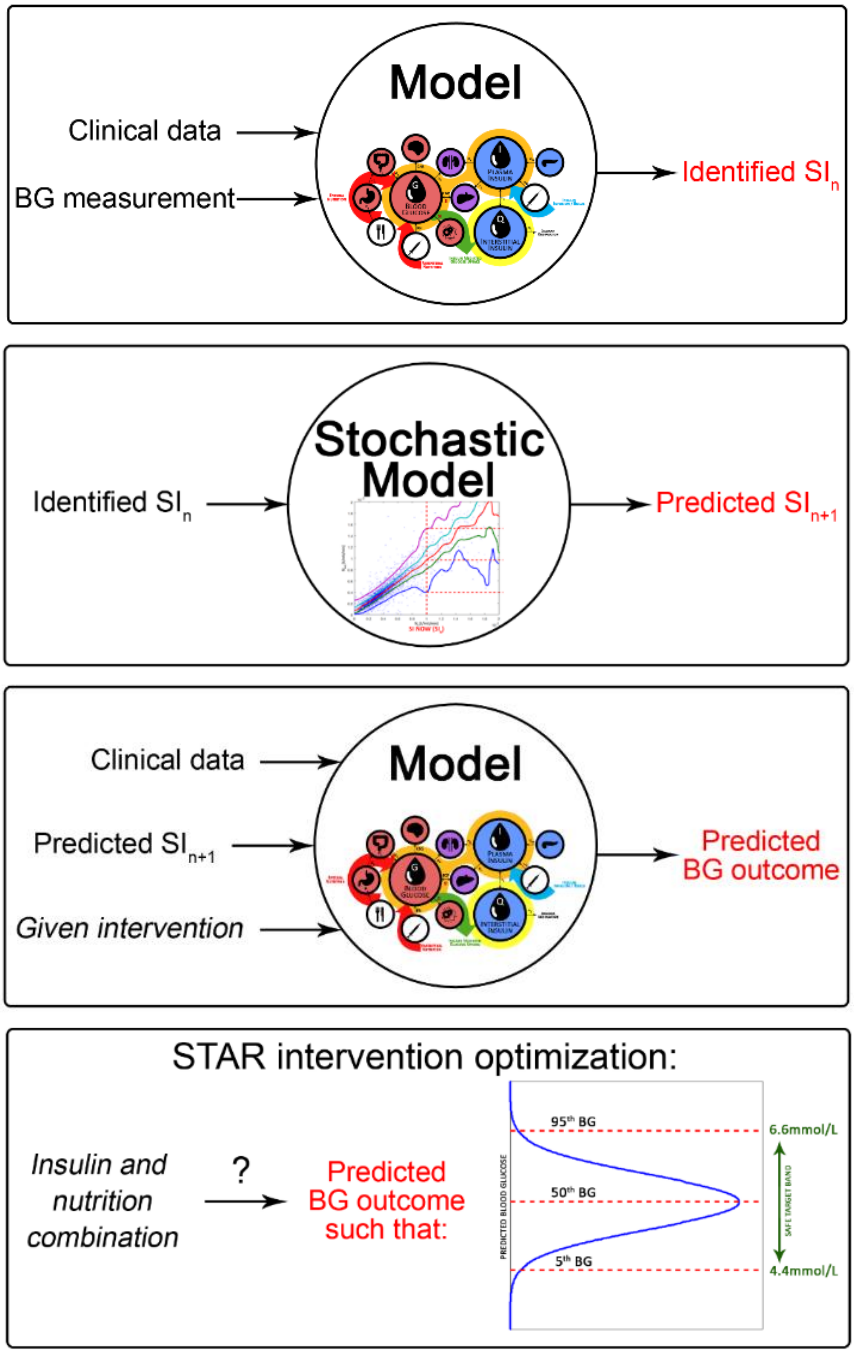

Figure 3.3 – STAR GC framework in 4 steps. Current SI is identified using the physiological model (top left). The stochastic model uses this value to predict future SI variability (top right). These predictions are used to calculate corresponding prediction of BG outcomes for a given treatment (bottom left). Treatment is adjusted to provide optimal control and minimise risks (bottom right).

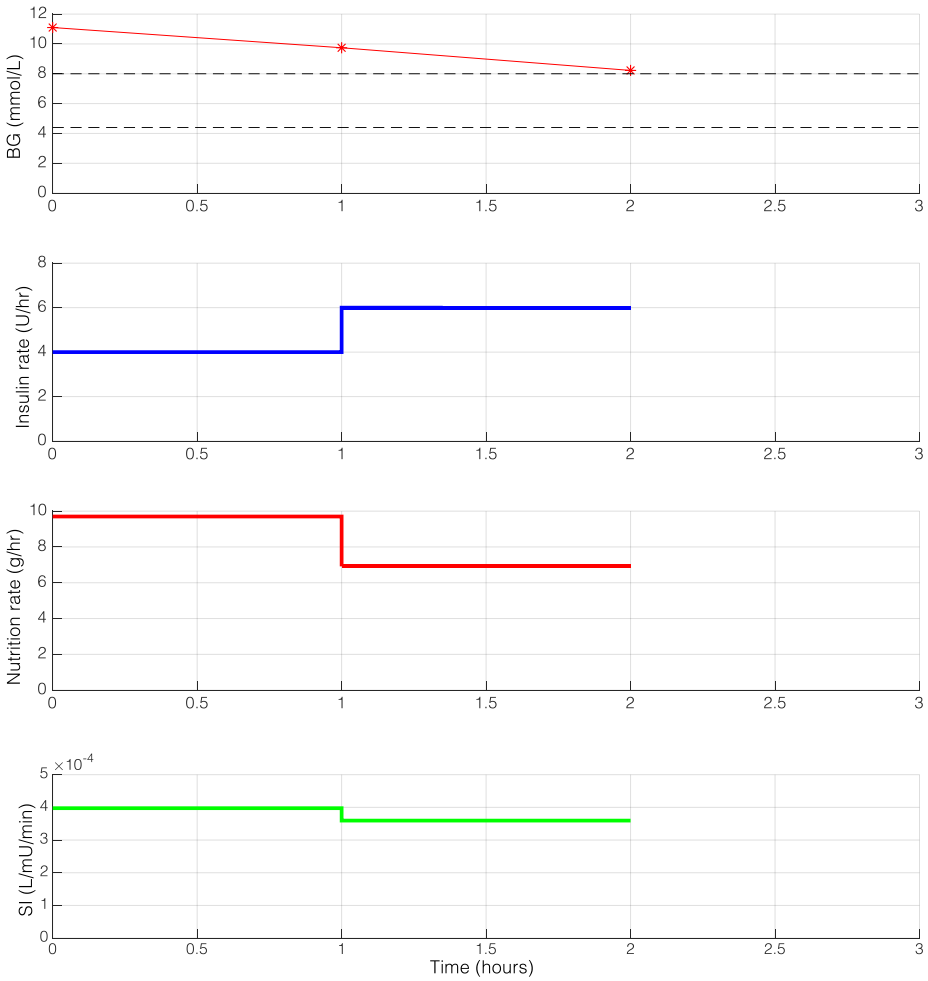

Figure 3.4 – Example of SI_n identification from clinical data. SI (bottom) is identified based on BG measurements (top), and insulin (second top) and nutrition (third top) rates over the identification period.

3.2.5.2. Step 2: Prediction of future SI variability

Given the current identified SI_n level, the stochastic model provides probability distribution functions of future likely SI_{n+1}, SI_{n+2}, and SI_{n+3}. The model is built from clinical data. More specifically, the 5^{th} and 95^{th} percentile prediction range is given by the stochastic model [124, 125]. These percentiles are directly used to optimise control, but, most importantly, to minimise the risk of hypoglycaemia in selecting the treatment. This step accounts for intra-patient variability, as it characterises the potential evolution of SI given its identified, patient-specific current value.

As shown in Figure 3.5, SI variability is typically larger as the time interval (1, 2, and 3 hours) increases. This increase reflects the intuitive higher potential variability when considering longer treatment intervals, reflective of likely higher risk of hypoglycaemia. Hence, it innately captures the trade-off of interval length (time) and potential variability.

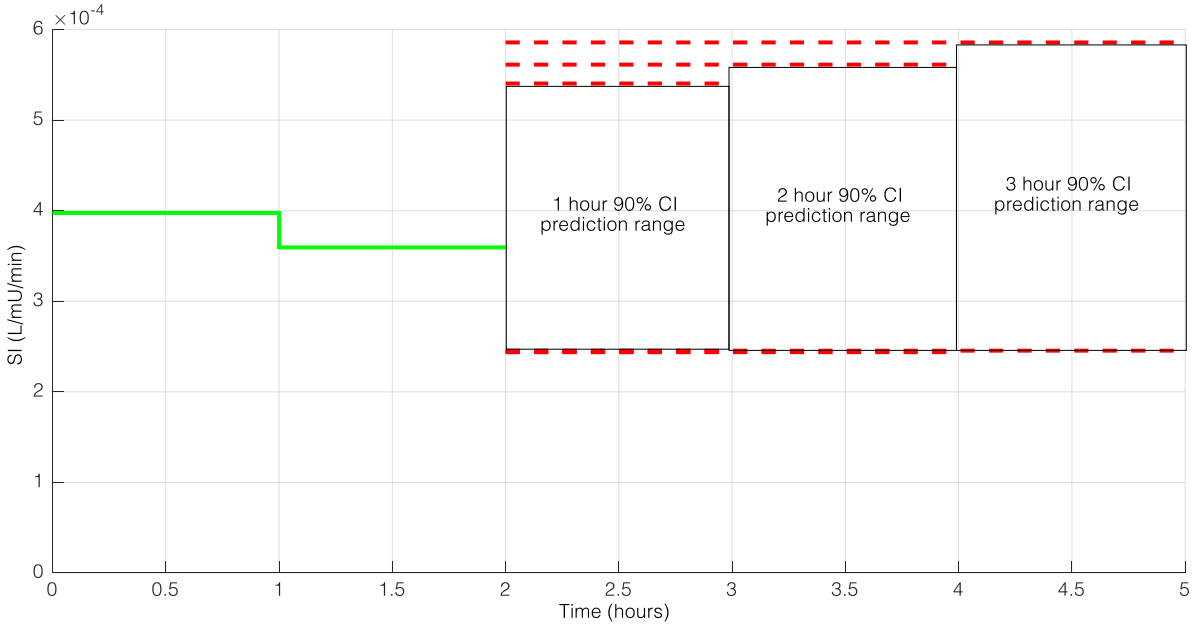

Figure 3.5 – Forward 1-3 hourly prediction of future SI variability. The 5th-95th percentile prediction ranges are determined by the stochastic model, based on the current identified SI value.

3.2.5.3. Step 3: From predicted SI to predicted BG

Based on the 5th and 95th percentile prediction range of future SI, a corresponding predicted BG outcome for any given combination of insulin and nutrition inputs can be calculated. This step uses the physiological model, but SI is now a known parameter with an expected range, and (future) BG is unknown. Figure 3.6 shows an example of the predicted 5th-95th percentile BG range (blue in top panel) based on the 5th-95th percentile range of predicted SI (dotted green bottom panel). This range was calculated for an insulin rate of 8U/h (second panel) and nutrition rate of 4 g/hr. As presented in Figure 3.7, this process can be done for all different, clinically acceptable treatment choices, impacting the evolution of predicted BG outcomes.

3.2.5.4. Step 4: Treatment Selection: Where the Magic Happens

Given likely future evolutions of BG levels for all acceptable treatments, STAR uses the 5th-95th percentile range in future BG outcomes to select the intervention that best overlaps the clinically specified target band [95]. Treatments are considered acceptable if the 5th percentile of future BG levels is above the lower limit of the clinically set target band. Typically, this target band is 4.4-8.0 mmol/L, and STAR thus limits the risk of BG < 4.4 mmol/L to a maximum of 5%. As a result, the risk of BG < 4.0 mmol/L is typically < 1%, with almost zero severe hypoglycaemic risk – providing the unique risk-based dosing approach.

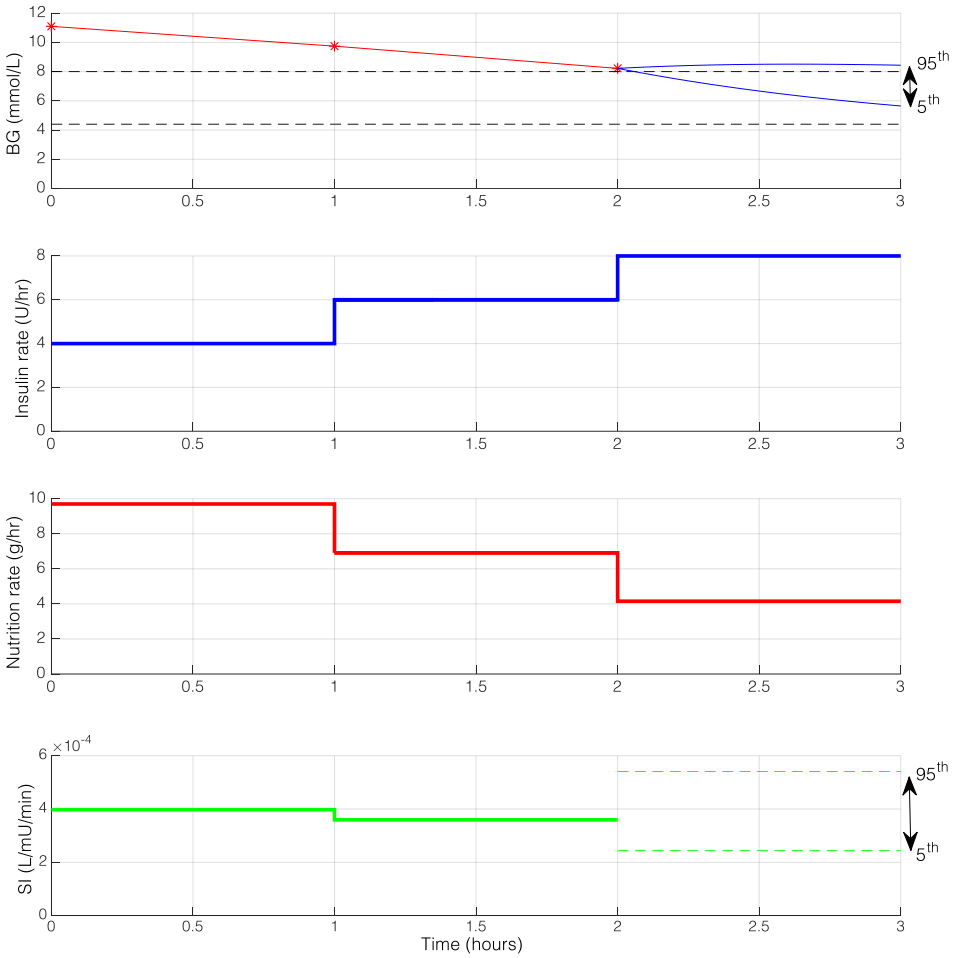

Figure 3.6 – The 1h forward 5th-95th percentile prediction range of BG (blue, top panel) calculated for the corresponding 5th-95th percentile prediction range of SI (dashed green, bottom panel), and specific insulin (second panel) and nutrition (third panel) inputs.

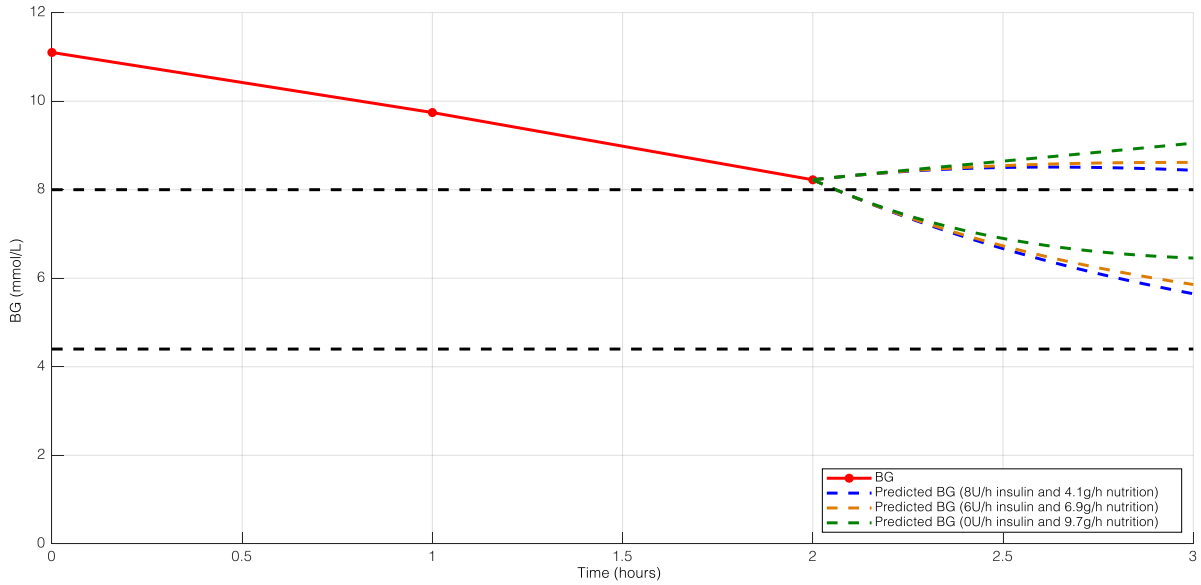

Figure 3.7 – 5th-95th percentile BG prediction ranges for different combinations of insulin and nutrition intervention.

In the overall example presented in this Section, the best 1-hourly treatment intervention (Figure 3.7) is the dotted blue line. This treatment is chosen as the associated 5th percentile predicted BG outcome is well within the target band, reflective of minimal risk of hypoglycaemia. Additionally, a lower nutrition rate (4.1g/h) is chosen compared to potential full maximum nutrition target rate (9.7g/h in this case), since the 95th percentile predicted BG is above the target band, thus reflective of potential hyperglycaemia. These choices can be modulated for different clinical practice cultures, of desired.

In Figure 3.8, the clinical BG evolution of this patient, receiving the treatment selected above, is shown over the following 1-hour period. The real BG evolution trace (red, top panel) is within the predicted range (blue, top panel). The identified SI for that period is shown (bottom panel), and also within the predicted SI range.

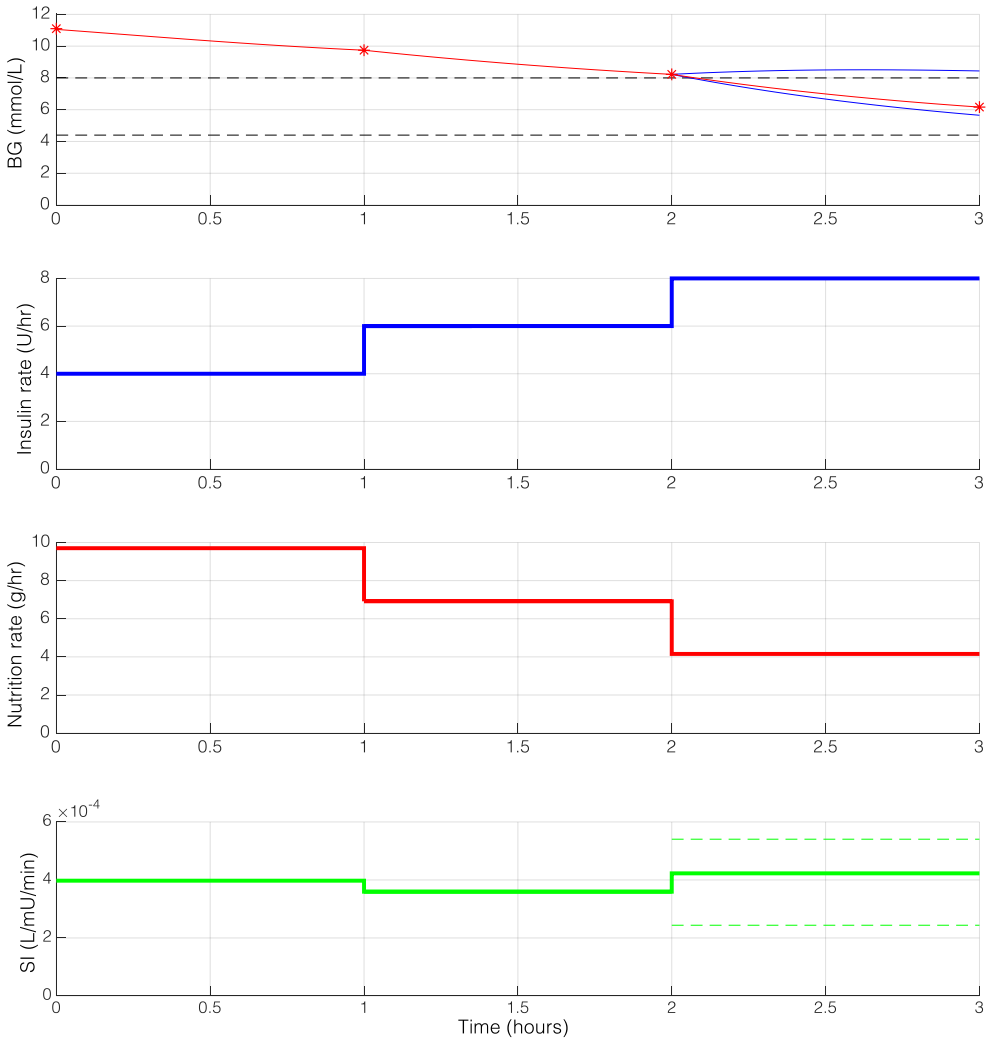

Figure 3.8 – Evolution of BG (top panel) given a specific insulin (second panel) and nutrition (third panel) inputs, and identification of the new SI level (bottom panel). Real BG (red) is well within the predicted range (blue). The new identified SI value (solid green) is also in the predicted range (dashed green).

While safety is the first priority, maximising nutrition to 100% of the patient-specific goal feed (GF) is the second [87, 134]. Nutrition can be decreased to a safe minimum value of 30% of GF [143] only if insulin is not sufficient to safely decrease BG outcomes within the target band. However, the nutrition rate can only be decreased by a maximum of 30% from one BG measurement to another to minimise over response to BG measurement errors. Insulin is limited to 6 U/h, with a potential additional 1-3 U/h of background continuous infusion for highly resistant patients [95]. This relatively low upper limit in comparison to other protocols has been set based on reported insulin-dependent receptor binding saturation [97, 132, 133]. When BG ≤ 3.0 mmol/L, insulin is stopped and a glucose bolus of 10 ml of 50% glucose (5g of glucose) is administered to the patient, while the next BG measurement interval is 1 hour.

Clinicians are free to adapt treatment according to their clinical judgment, but these changes are recorded and accounted for in future treatment calculations. Insulin can be administered as continuous intravenous infusions or hourly boluses, showing STAR's ability to adapt to local standards. Treatments are re-evaluated by the STAR controller at each BG measurement, where measurements are taken using either glucometers or blood gas analyser. STAR protocol treatments are calculated for 1, 2, and 3-hour treatment intervals, where the 2 and 3-hour treatment options are offered depending on glycaemic stability within the targeted range, and safety from hypoglycaemia. Thus, BG measurements are taken 1-3 hourly.

3.2.6. Safety and Efficacy of the STAR Framework

Safety and performance analyses on STAR have shown its ability to provide safe and effective control for ICU patients in two hospitals from New Zealand and Hungary [87]. Over 250 patients (336 GC episodes) at the Christchurch Hospital, New Zealand, and 47 patients at the Gyula Hospital, Hungary, treated under STAR were analysed in this study. In each arm, the percentage time of BG in the target band (4.4-8.0 mmol/L) was > 80%, and the incidence of hypoglycaemia was very low, with < 2%BG below the lower limit of the target. Based on these encouraging single centre results, each hospital adopted STAR as the standard of care for GC.

While STAR has shown promising results, it has only been implemented in single centre clinical trials, rather than expensive, complex to implement and compare, international multi-centre randomised

clinical trials (RCTs). The statistical power of the STAR results is thus often underestimated, although there are debates on whether international multi-centre RCTs are statistically more powerful, due to their potentially large variability and low compliance across ICU settings and practices [144-148]. This issue is further examined in Chapter 5.

It is also important to remember there are some limitations in the GC controller arising directly from its design. STAR uses a physiological model, where some parameters have been approximated or identified from literature. For example, it is very complicated to assess the exact real-time evolution of EGP or insulin secretion rates. There thus might be some sources of error in the model, although the model has been validated and used extensively [87, 90, 91, 93, 94, 119, 149-151]. Additionally, the stochastic model is based on population data, and some specific behaviour in future variability might not be completely characterised.

3.3. Virtual Trials and Virtual Patients

Virtual trials are a powerful tool to compare glycaemic outcome, performance, and safety of different protocols simulated on the same set of virtual patients [91, 98, 99]. It can thus be used to analyse and compare how specific patients, characterised by selected key physiological parameters would react to different treatment approaches. The emergence of model-based solutions has led to the use of such *in-silico* trials [98].

More specifically, the mathematical model and GC protocol design can be tested, adapted, and validated using simulations of virtual trials on virtual patients. Protocols are thus not developed through trial-and-error methods, but rather optimised prior clinical implementation [99]. This allows simulations to potentially identify undesired negative effects from protocol design. It thus reduces the costs associated with multiple pilot clinical trials, and potential avoidable risk [91].

It is important to note virtual patients and virtual trials simulations imply perfect clinical compliance to protocol, thus reflecting a zero-error case. Results will be biased in clinical practice if compliance to protocol is not fully respected, or if significant clinical errors arise due to protocol complexity [152]. They will also be inaccurate if the virtual patient model is not fully validated [91].

In the context of this thesis, virtual patients are characterised by their unique metabolic model-based SI profiles, identified from clinical data. The ICING model used is fully validated and clinically well used [81, 93, 95, 99, 137, 149, 153-156]. For each patient, SI is identified hourly from clinical data. Importantly, patient-specific SI traces have been shown independent of the clinical inputs used to generate them [94], characterising thus patient metabolic state evolution over time [128]. Different treatment approach can thus be tested based on protocol specification, as shown in Figure 3.7.

It is important to mention virtual trials of STAR automatically select the longest treatment intervals suggested and available. Thus, if only 1-hourly is suggested by STAR for safety reasons, virtual trials will select this treatment. However, if STAR assessment of risks results in allowing 2, or 3 hourly treatment intervals, than the longest available will be automatically selected. These trials are thus "blind" to any other potential factors, such as low BG levels, that could affect nurse treatment selection in clinical use. Additionally, the ICING model used enables to simulate a protocol using exogenous insulin infusion, insulin boluses, or both, based on protocol design or ICU practices.

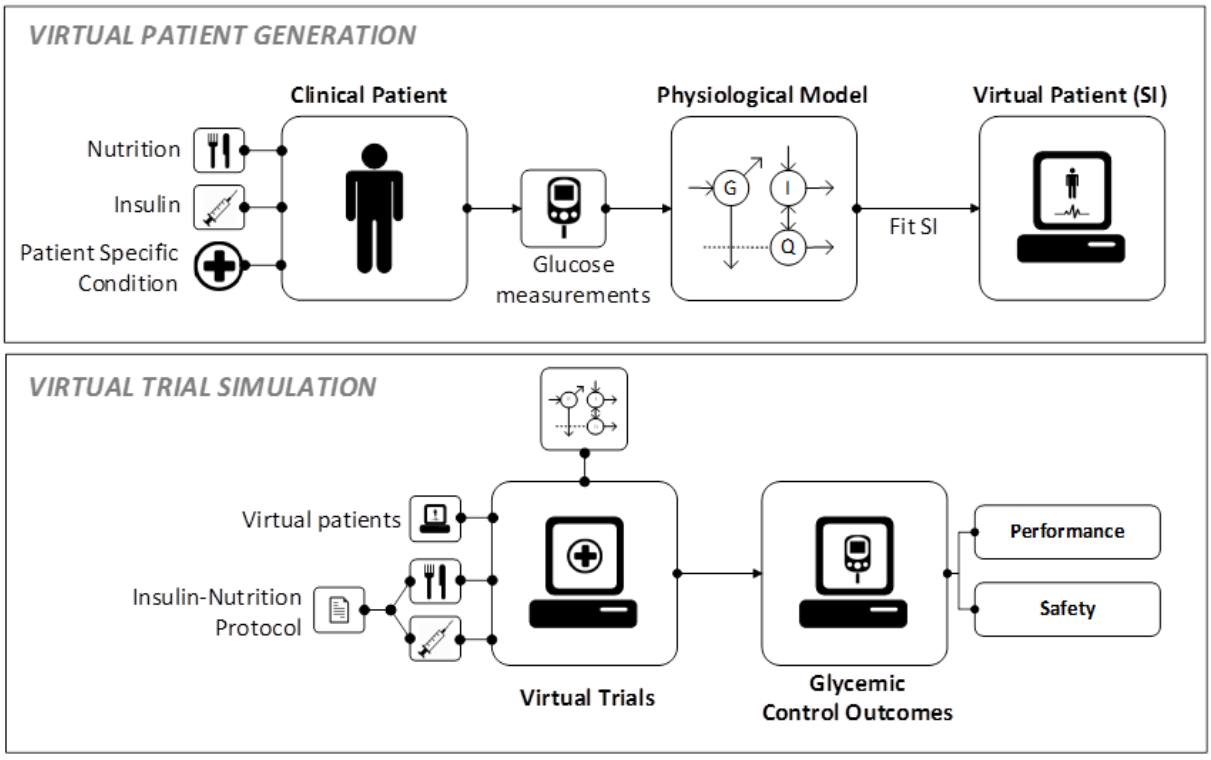

Figure 3.9 – Schematic representation of virtual patient generation and virtual trial simulation methodology. Virtual patients are characterised by their hourly identified SI from clinical data. Different protocol can then be tested on these patients to assess simulated BG outcomes in individual virtual patients and across a cohort. Figure taken from [128].

Such trials are well-validated in their independence from the data used to create them and their accuracy [91, 94, 128], their ability to predict trial outcomes [96, 120, 129] and in clinical use to guide care in STAR [87, 95, 134]. These virtual trials have been used to assess and optimise STAR in both adult and neonatal ICUs [94, 95, 105, 128, 129, 149, 151, 154, 157, 158]. The overall schematic illustration of these *in-silico* trials using virtual patients is presented in Figure 3.9.

3.4. Summary

This chapter presented and detailed the proven STAR GC framework, as well as the use virtual trials and virtual patients. Both tools will be used in this work to respond to the main questions and objectives of this thesis, as presented in Chapter 1. The STAR model-based GC framework uses a physiological and stochastic model to provide a unique risk-based dosing approach. It has been shown safe and effective in different ICUs and practices. Virtual trials have proven effective in assessing and comparing the outcomes of different GC protocols on the same underlying patients or cohort of patients.

Chapter 4: Cohorts, Patients, and Episodes

The retrospective analysis in this thesis are mainly using patient data from 3 different cohorts of patients, from 3 independent studies. These 3 cohorts are presented in this chapter, and have been previously published in [87].

4.1. GC Patients vs. Episodes

It is first important to note a patient can have multiple GC episodes, generally because:

- Patients BG is stabilized, but then several hours later GC is started again due to dysglycaemia arising from any potential clinical reason, or
- Patient are sent out if the ICU for clinical procedures (most commonly imaging or surgery), where GC is stopped and started again as they return (if necessary).

4.2. SPRINT Cohort (Christchurch)

This SPRINT protocol is the precursor of STAR, and provided safe, effective control for nearly all patients, averaging 16 measurements per day [38, 81, 87, 106]. SPRINT is a table-based protocol designed to adapt and optimise insulin and nutrition using BG assays and previous intervention. It was developed using the ICING physiological model, but does not explicitly identify SI. The BG target band in SPRINT BG is 4.0-6.1 mmol/L. A significant advantage of this data set is it contains a full set of clinical data of sufficient detail and quality for a wide range of further analyses. Many studies do not record (or report) detailed nutrition and/or insulin inputs, and so limit analyses by either disregarding nutrition in the first place, or by only reporting daily averages and effects. This data set included all time-valued changes in insulin and nutrition in 1-3 hour intervals, as well as all BG measures, thus allowing a much higher degree of resolution in the calculation of time-varying SI.

In total, 371 patients of the SPRINT study [81, 94] are considered. These patients were treated in Christchurch Hospital ICU, New Zealand, from July 2005 to May 2007. SPRINT was implemented as standard practice, and de-identified data audit and analysis were approved by the New Zealand Health and Disability Ethics Committee Upper South Regional Ethics Committee B (**Ref: URB/07/15/EXP**).

From this cohort, 292 (79%) patients with GC episodes longer than 10 hours and average nutrition < 120%GF are used. In addition, episodes were split if a gap greater than 5 hours between two consecutives measurements was present, leading to a total 442 GC episodes. These criteria ensure the normal use of SPRINT and the exclusion of patient data with very short GC episodes, and thus low BG measurement numbers, which are likely less reflective of general metabolism dynamics [87]. Demographics summary is shown in Table 4.1.

Table 4.1 – Summary of SPRINT patient demographics.

	SPRINT Christchurch
# episodes	442
# patients	292
# GC hours	39838
% male	62.7
Age (years)	63 [48, 73]
APACHE II	19.0 [15.0:24.5]
ICU Length of stay (days)	6.2 [2.7,13.0]

Results are given as median [IQR] where appropriate.

4.3. STAR Cohorts (Christchurch and Gyula)

Retrospective patient data included on STAR from 2 clinical ICUs are also used. The first cohort includes 264 patients treated in Christchurch Hospital ICU, Christchurch, New Zealand, from June 2011 to May 2015. The second cohort includes 47 patients treated in Kalman Pandy Hospital ICU, Gyula, Hungary, from December 2011 to May 2015. STAR in both countries were implemented as standard practice, and de-identified data audit and analysis were approved by the New Zealand Health and Disability Ethics Committee Upper South Regional Ethics Committee B (**Ref: URB/07/15/EXP**), and the local ethical codes of Hungary.

Compared to the SPRINT patients, the data is sparser as STAR averages 12 measurements per day. However, BG, insulin, and nutrition rates recoding were automated based on clinical staff entries, and not gathered from patient bed sheets. Additionally, these patients' lower measurement frequency compared to SPRINT is closer to most ICUs standards.

These cohorts are totalling 330 and 47 GC episodes for STAR-Christchurch and STAR-Gyula, respectively. Similarly to the SPRINT cohort, only episodes longer than 10 hours and average nutrition < 120%GF are considered. Demographics results summary are presented in Table 4.2.

STAR in these two different cohorts was implemented slightly differently. In Christchurch, insulin boluses are primarily used, and enteral nutrition is modulated. In contrast, in Gyula, insulin infusion is mainly

used, and parenteral nutrition is modulated. Thus, these two cohorts, in addition of being different ethnically, are representative of different GC management in the context of STAR.

Table 4.2 –Demographics summary of STAR Christchurch and STAR Gyula cohorts.

	STAR Christchurch	STAR Gyula
# episodes	330	47
# patients	264	47
# GC hours	22372	6268
% male	65.5	61.7
Age (years)	64 [53, 72]	66 [58, 71]
APACHE II	21.0 [16.0:25.0]	32.0 [28.0:36.0]
LOS - ICU (days)	5.7 [2.5,13.4]	14.0 [8.0,20.5]

Results are given as median [IQR] where appropriate.

4.4. Summary

This chapter presented the retrospective cohorts used in analyses of this thesis. These cohorts reflect different population of patients, treated with different GC protocols proven to achieve highly effective GC. In addition, the quality of the data is an advantage to accurately identify patient-specific SI and create virtual patients (Chapter 3).

Chapter 5: Bad vs. Good Glycaemic Control

The disparities in clinical results of tight GC, targeting lower BG targets independently associated with improved outcome, resulted in ongoing debate on the optimal target band. More specifically, while some studies have shown improved outcome providing safe control, others resulted in significant increased hypoglycaemic risk, associated with worse outcome.

This chapter aims to understand and identify why some studies were successful in providing safe GC control and others were not. More specifically, it provides a first answer to the question: Is GC to lower ranges to blame for increased hypoglycaemia and poor results, or rather the control given? This goal is achieved by comparing clinical results and validated virtual trial results of two contradictory studies to assess the cause of their differences in outcome.

This chapter presents results published in [159] and [160].

5.1. Introduction

Stress-induced hyperglycaemia is a common complication in ICU [31], associated with increased morbidity and mortality [5, 13]. GC to lower BG to safer ranges have associated with improved outcomes [35, 161, 162], but hard to achieve safely [46-48, 50, 89], increasing hypoglycaemia and glycaemic variability, both associated with worsen outcome [5, 10, 11, 13, 14, 32, 33, 54, 56, 58, 163]. Whether insulin therapy to provide GC is beneficial or harmful for critically ill patients has thus been widely debated over the last 20 years [60, 61, 63-68, 164-168]. More precisely, while GC is well accepted, the optimal BG target band is still unknown [121]. High time in normoglycaemic ranges is associated improved outcome [38, 76, 77, 79], but is harder to achieve safely due to metabolic variability [79, 85]. To date, the recommendations advocate moderate, rather than tight, BG targets for GC in ICU [70-73], reflecting the “first do no harm” or “uncertainty” principle [60]. However, these recommendations are heavily based on studies that failed to provide safe and effective control for all patients [75].

Thus, it is unclear why some studies were able to provide “good” or “bad” control. Understanding what makes GC hard to achieve safely could thus provide a first answer to this question: Is intensive insulin therapy (IIT) for tight GC to blame for increased hypoglycaemia and poor results, or is it the control given?

5.2. The Bad vs The Good

Chapter 1 presented the importance of model-based GC protocol, accounting for inter- and intra- patient variability. Effectively, automated GC protocols allow identification of patient-specific key physiological parameters reflecting patient metabolic state, such as SI [85], and thus the ability to adapt treatment accordingly. Compliance to protocol and measurement frequency also directly impact GC performance [169, 170], and are added key considerations, as they reduce controller performance. Thus, effective protocol must consider much more beyond metabolism to ensure best outcomes.

In 2009, the Normoglycemia in Intensive Care Evaluation-Survival Using Glucose Algorithm Regulation (NICE-SUGAR) study results – a large, important multi-centre RCT comparing conventional and intensive GC impact on 90-day clinical outcomes – were presented and quickly became the gold standard in the field [46]. This study showed the important increased risk of severe hypoglycaemia and

mortality when using the lower target band compared to a higher glycaemic target with their protocol [32, 46]. However, the methods, results, and lack of reported statistics have been widely criticised [63, 121, 164], making the conclusion of this study about 90 day mortality questionable. Overall, this study created a “valley of glycaemic despair” [64], where experts begged to “agree to disagree” [61], and the field has since languished.

A decade has passed since the NICE-SUGAR study, and new studies have once again shown improved control and achieving clinical outcomes are becoming (potentially) possible using lower glycaemic targets [43, 87, 95, 108, 149, 151, 171, 172]. Unlike NICE-SUGAR, all these studies used computerised methods and achieved high time in the targeted band in a safe, effective manner. Unfortunately, their statistical power, size, and/or retrospective nature is criticised compared to large RCTs.

Whether international multi-centre RCTs are statistically more powerful, despite their potentially large variability and low compliance across ICU settings and practices, is also hotly debated [144-148]. In particular, no clinical study can compare two or more protocols on the same patient, so there is no way to determine if poor RCT results are due to the protocol, compliance, or GC itself. Additionally, the lack of standardised reported metrics and statistics have been pointed out as a big problem in studies analysing GC outcomes, making them hard to compare fairly (Chapter 2) [99, 118].

Virtual trials can be used to test multiple protocols on the same cohort of virtual patients. These *in-silico* trials were further explained in Chapter 3. This analysis uses clinically validated virtual patients [91, 94, 128] to simulate the NICE-SUGAR protocol intensive low target arm, and assesses its ability to achieve safe, effective control. Results are compared with the reported clinical data. Simulated glycaemic outcomes of this protocol are also compared with virtual trials of the proven, model-based and patient-specific STAR GC framework [87, 95] using the same underlying virtual patients.

Overall, this control performance analysis aims to determine whether GC has been the scapegoat of the NICE-SUGAR study, wrongly blamed for poor patient outcomes due instead to the ineffectiveness of GC design.

5.3. Methods

5.3.1. Glycaemic Control Protocols

5.3.1.1. STAR Protocol

STAR has been fully described in Chapter 3, and is used in this analysis. STAR is a model-based GC framework used as the standard of care in adult ICU of Christchurch, New Zealand, and Gyula, Hungary since 2012 [87, 95, 123]. STAR provides safe, effective control for effectively all patients to lower glycaemic ranges (4.4-8.0 mmol/L), using a unique risk-based dosing approach [95, 125], modulating both insulin and enteral nutrition inputs.

5.3.1.2. NICE-SUGAR Protocol

The NICE-SUGAR study was an international multi-centre RCT comparing intensive and conventional glucose control [46]. The intensive control group target was 4.5-6.0 mmol/L, while the conventional group was 10.0 mmol/L or less [46]. Intravenous insulin infusion was used alongside glucose boluses, and caloric nutrition intake was carried out at clinician discretion and was not considered in the protocol, making it “CHO blind” [85, 86]. BG measurements were made using local ICU standards.

The full protocol algorithm can be found online [173], and corresponds to a sliding-scale protocol using current and previous BG measurements, and current insulin dose to adapt treatment. By protocol design, measurements are required to be taken 1-hourly, and 2-hourly if BG is within target band and insulin infusion and caloric intake are stable. If the last 2 BG measurements are in band and no insulin is given, the measurement frequency is 4-hourly. Finally, 30-minute measurements are required if BG is too low (BG < 4.5 mmol/L). The workload is thus expected to be important as the measurement frequency is high: primarily 1-2-hourly.

Although it may seem unreasonable, no maximum recommended insulin infusion rate information was found in the published study protocol. Insulin recommendations may be adapted at clinician discretion if necessary, but no specific guidelines are given. Insulin infusion is stopped and glucose boluses of 10 ml and 20 ml of 50% glucose are given if BG drops below 3.5 mmol/L and 2.5 mmol/L, respectively. Equally, nutrition was left to local guidelines and practice, and was not specified by the protocol.

5.3.2. Virtual Trial Analysis

Validated virtual trial using virtual patients (Chapter 3) is used to simulate different GC protocols and compared GC outcome from the same underlying cohort of patients [91, 94, 128]. In this analysis, 442 virtual patient episodes are generated using retrospective clinical data from 292 patients of the SPRINT cohort presented in Chapter 4 [81, 94]. Three different protocols are simulated in this study: the intensive NICE-SUGAR protocol (NS-IIT); a 3-hourly measurement interval adapted version of the intensive NICE-SUGAR protocol (NS-IIT-3H); and STAR.

A summary of these protocol design is presented in Table 5.1. While NS-IIT strictly implements the published protocol, NS-IIT-3H is developed to average 3-hourly measurements and limit insulin infusions to a maximum of 10 U/h. This modified NICE-SUGAR protocol was developed to try to match published clinical practice results and are also a better comparison in terms of workload to STAR. The NS-IIT-3H version also reflects a more manageable and realistic GC design, directly matching the reported clinical workload, which was much lower than the 1-2 hourly interval specified in its original protocol design. Overall, these three analyses enable a fair, more realistic comparison with both reported clinical results and STAR. More specifically, measurements for NS-IIT-3H were changed from 1 to 3-hourly (for the same insulin treatment dose) if within 4.5-10.0 mmol/L, and maintained unchanged at 1-hourly (or 0.5-hourly) for lower BG values as per original protocol.

Table 5.1 – Summary of the NS-IIT, NS-IIT-3H, and STAR protocol designs compared in this analysis.

	NS-IIT	**NS-IIT-3H**	**STAR**
Type	Table-based	Table-based	Model-based
Target band	4.5-6.0 mmol/L	4.5-6.0 mmol/L	4.4-8.0mmol/L
Measurement intervals	0.5 hourly if low BG 1-2 hourly if in band 4-hourly if stable.	0.5 hourly if low BG 3-hourly otherwise	1-3 hourly
Insulin strategy	No limit specified	Max. 10 U/h	Max 6U/h + 3U/h if persistent hyperglycaemia
Nutrition strategy	Clinical discretion	Clinical discretion	Modulation between 30 to 100% GF

In the simulation of the intensive NICE-SUGAR protocol (NS-IIT), nutrition intake is increased daily, according to reported non-protein calorie achievements in the Appendix B of the original study [46]. However, nutrition type and composition are never reported in the study. Therefore, enteral nutrition is assumed to be low CHO (40% of non-protein calories were CHO, similar to Glucerna 1.0™), and parenteral (including maintenance fluids) are 70% of non-protein calories attributable to CHO. This low CHO enteral nutrition assumption is conservative and provides a best case for the NICE-SUGAR protocol, considering the higher CHO content typical of most enteral nutrition formulas would make GC more difficult, increasing glycaemia and patient variability.

5.3.3. Safety, Performance, and Compliance Assessment

Many statistics exist to assess GC performance, and all have advantages and disadvantages or limitations [99, 116-119]. These issues are further discussed in Chapter 2. However, in this analysis, performance and safety metrics are chosen based on the few BG statistics reported in the NICE-SUGAR study and more commonly used metrics in the literature. Thus, performance is assessed by the mean and SD of hourly resampled BG levels, percentage time of hourly resampled BG measurements within the different target bands (4.5-6.0 mmol/L for NICE-SUGAR and 4.4-8.0 mmol/L for STAR), as well as median, inter-quartile range and 5-95th percentile range of per-patient mean BG. Safety is evaluated by the number of patients and total %BG in severe hypoglycaemia (BG < 2.2 mmol/L). Mild hypoglycaemia is also reported for each protocol (%BG < 4.0 mmol/L). Finally, workload (average measurements per day), mean insulin rate and median glucose rate are reported, where major differences in workload compared to clinical results indicate non-compliance to the protocol.

As the NICE-SUGAR study did not report full results, comparison of BG metrics with the simulations is more difficult. Hence, only the clinical mean (SD) BG is reported in the results comparison. Additionally, the time-weighted per-patient BG is derived from Figure 2 presented in the original study [46]. To ensure strong and robust comparison, and because the time-weighted per-patient BG is already an estimation of presented results, no additional estimation on per-patient BG metrics is made because there would be too much room for error. Such lack of standardised reporting metrics for GC in ICU has been addressed and reported previously [90, 99, 118], but unfortunately is still today too often an issue.

5.4. Results

Virtual trial and clinically reported results for NS-IIT, NS-IIT-3H, and STAR are presented in Table 5.2. Per-protocol CDFs of hourly resampled BG and per-patient mean BG are shown in Figure 5.1. Per-protocol median [IQR] evolution of hourly resampled BG overtime stratified in 6-hour blocks is also shown in Figure 5.2. Clinical results for the NICE-SUGAR study are not shown in Figure 5.2 and some parts of Table 5.2 because not enough data was reported to reconstruct or estimate BG outcome ranges.

5.4.1. NS-IIT vs. STAR Virtual Trial Results

STAR outperforms NS-IIT in safety and GC performance. Mean BG was similar (6.2±1.2 vs. 6.2±1.7 mmol/L), and both inter- and intra- patient glycaemic variability were lower for STAR as reflected in the median [IQR] per-patient mean BG (6.2 [5.9, 6.6] vs. 6.2 [5.7, 7.1] mmol/L) and SD BG (1.4 [1.1, 1.8] vs. 1.1 [0.8, 1.4]). Figure 5.1 and Figure 5.2 also show STAR has consistently tighter and less variable BG outcomes compared to NS-IIT. STAR achieved 91% BG within 4.4-8.0 mmol/L (81% for the NS-IIT), and both protocols achieved similar 46% BG within 4.5-6.0 mmol/L. In terms of safety, BG < 4.0 mmol/L was only 1% mild hypoglycaemia and 0.02% in severe hypoglycaemia for STAR compared to 3% and 0.04% for NS-IIT. In total, 5 patients (1%) had sever hypoglycaemic episodes with STAR and 7 (2%) with NS-IIT. Additionally, 55% of patients experienced mild hypoglycaemia (BG within 2.3-3.9 mmol/L) under the NS-IIT protocol compared to 34% for STAR.

Considering compliance, workload was, as expected, much higher for NS-IIT (25.0 measurements per day) compared to STAR (12 measurements per day). Insulin rates were much higher for NS-IIT (154.6±209.2 U/d) compared to STAR (70.4±53.5 U/d). While this higher insulin rate could be explained by the higher median CHO intake (6.2 [5.1, 6.7] vs. 6.1 [2.1, 6.1] g/h), the absence of a maximum insulin limit also plays a role, using insulin in an ineffective and potentially dangerous manner, as seen in the higher hypoglycaemia reported. Finally, the 5-95th percentile range of per-patient mean BG indicates 5% of patients were above 9.7 mmol/L under NS-IIT compared to 8.1 mmol/L with STAR.

Table 5.2 – Simulation results summary of the NS-IIT, NS-IIT-3H, and STAR protocols, and reported clinical values of the NICE-SUGAR study.

	NICE-SUGAR			STAR
	Clinical	NS-IIT	NS-IIT-3H	Simulation
Average measurements per day	~9.4	~25.0	~10.2	~12.0
Mean insulin dose (SD) U/day	50.2 (38.1)	154.0 (209.2)	115.0 (121.7)	70.4***** (53.5)
Mean resampled BG (SD) [mmol/L]	6.4 (1.0)	6.2 (1.7)	6.4 (1.7)	6.2* (1.2)
Median [IQR] per-patient mean BG [mmol/L]	/	6.2 [5.7, 7.1]	6.4 [5.9, 7.3]	6.2† [5.9, 6.6]
5-95th percentile per-patient mean BG [mmol/L]	/	[5.3, 9.7]	[5.3, 9.6]	[5.4, 8.1]
Median [IQR] per-patient SD BG [mmol/L]	/	1.4 [1.1, 1.8]	1.3 [1.0, 1.7]	1.1† [0.8, 1.4]
5-95th percentile per-patient SD BG [mmol/L]	/	[0.8, 2.7]	[0.6, 2.6]	[0.6, 2.4]
% BG in 4.5-6.0 [mmol/L]	/	46	41	46‡
% BG in 4.4-8.0 [mmol/L]	/	81	81	91‡
% BG < 4.0 [mmol/L]	/	3	2.5	1‡
% BG < 2.2 [mmol/L]	/	0.02	0.06	0.02‡
# patients with min(BG) within 2.3-3.9 mmol/L (%)	2237 (74)	243 (55)	210 (47)	149‡ (34)
# patients with min(BG) <= 2.2 mmol/L (%)	207 (7)	7 (2)	20 (5)	5‡ (1)
Median [IQR] glucose rate (g/h)	/	6.2 [5.1, 6.7]	6.2 [5.1, 6.7]	6.1† [2.1, 6.1]

Simulation BG data is resampled hourly. "/" is used if no data is available in the original study. Significance level (P<0.05) is indicated with "" for Welch Test, "†" for the Wilcoxon rank sum test, and "‡" for the Fisher exact test, comparing STAR with NS-IIT-3H. SD is standard deviation.*

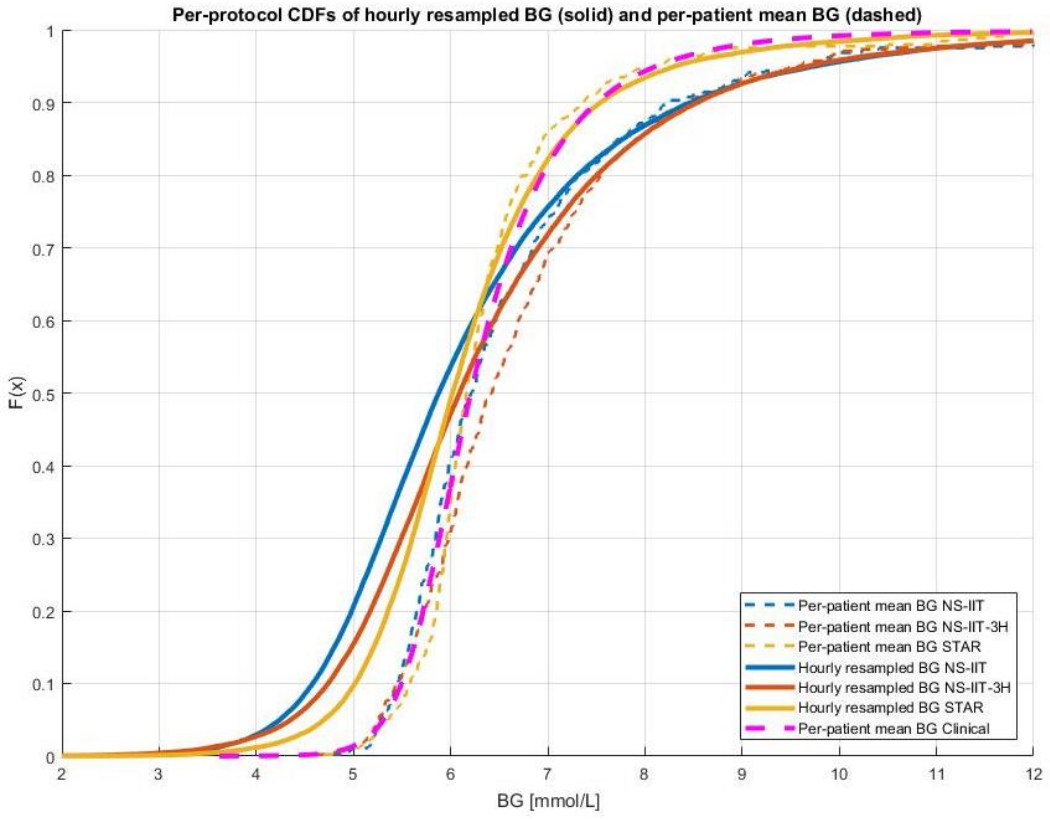

Figure 5.1 – Per-protocol cumulative distribution function of BG. Solid lines represent hourly resampled BG. Dotted lines represent per-patient mean BG. Clinical CDF is estimated from [46].

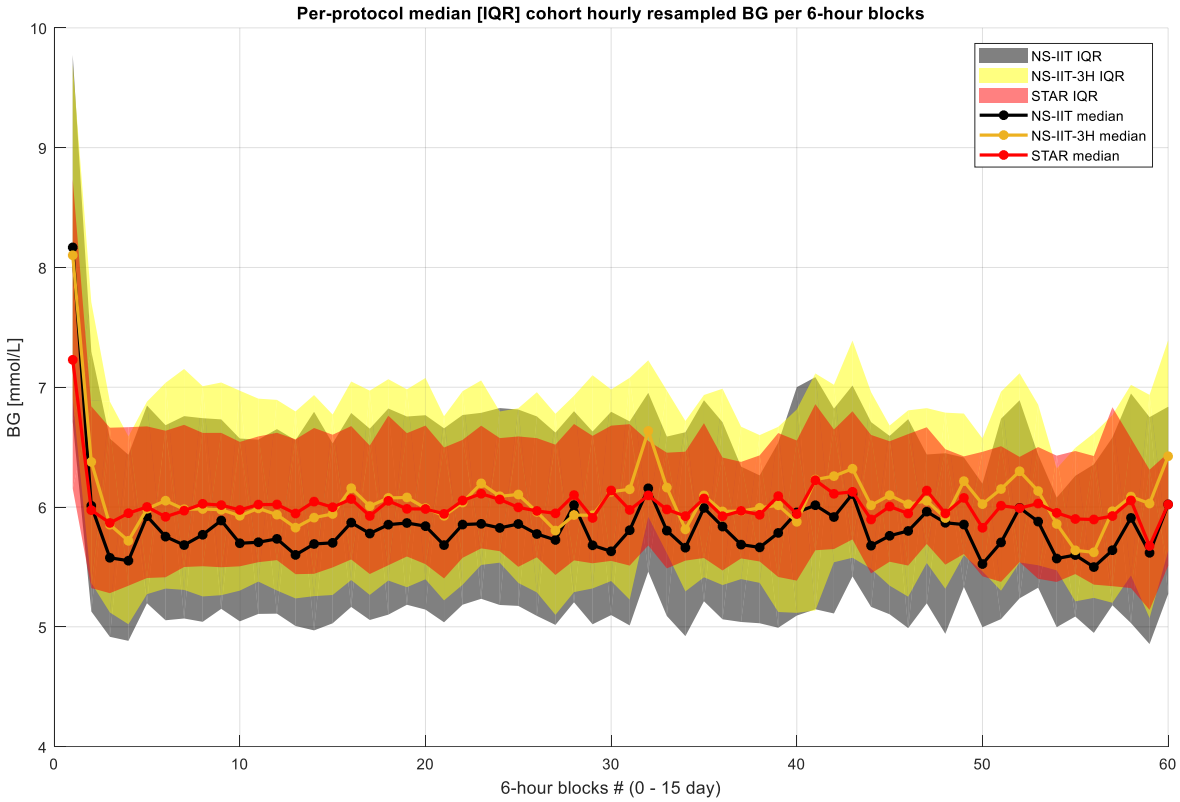

Figure 5.2 – Per-protocol median [IQR] evolution of hourly resampled BG stratified in 6-hour blocks over the first 15 days of control. Solid lines and dots indicate 6-hourly medians and shaded areas represent IQR.

5.4.2. NS-IIT Clinical vs. Virtual Trial Results

Reported clinical results of the NS-IIT study differ significantly from simulations. Although the mean resampled BG levels are similar (6.4±1.0 vs. 6.2±1.7 mmol/L), the number of patients experiencing severe and mild hypoglycaemia is larger clinically (6 vs 2% and 74 vs. 55%). This difference could be explained by the large difference in workload (9.4 vs. 25.0 measurements per day), clearly showing a clinical lack of compliance to the original NS-IIT protocol, whose measurement rate was most likely not clinically feasible [81, 174, 175].

In addition, the mean insulin rate clinically was only 50.2±38.1 U/d, while exact per protocol virtual trial simulation results show much higher rates (154.0±209.2 U/d). Once again, this difference with clinically reported data is reflective of poor compliance to protocol, particularly as insulin levels commanded per protocol increased. Recommendations of excessive, ineffective amounts of insulin were probably lowered at clinician or nurse discretion for safety or out of fear of the risk of hypoglycaemia, but the degree of non-compliance is not clear nor reported in the original study [46], and may not have been recorded.

Another explanation could also be the underlying cohort in the original study was less resistant to insulin than the virtual cohort or were fed less than in simulations. Importantly, the virtual trial cohort used from Christchurch, New Zealand, is demographically and clinically similar to many of those in the NICE-SUGAR trial in Australia, New Zealand, and Canada. Any or all of these factors may play a role. However, with such a large peak insulin dose, and very similar cohort, as well as other large differences, it would not be surprising to find compliance to such recommendations to be low [176-178].

5.4.3. NS-IIT-3H

NS-IIT-3H was implemented to better match average reported clinical measurements and allow fair comparison with STAR based on workload, as clinically reported. With 10.2 measurements per day, NS-IIT-3H better matched the 9.4 measurement per day reported clinically, better capturing reported workload. Results show 81% of NS-IIT-3H BG values within the STAR target band and 41% within the NS-IIT target band. The mean BG level was higher than STAR (6.4±1.7 vs. 6.2±1.2 mmol/L), but very close to reported values (6.4±1.0 mmol/L).

Interestingly, the NS-IIT-3H per-patient mean BG CDF is very different from the estimated clinically reported results, the latter being closer to STAR (Figure 5.1). This result indicates the compliance to protocol is more than just increased measurement intervals. Figure 5.2 shows the NS-IIT-3H protocol design shifts the median [IQR] BG upwards, bringing the median BG closer to STAR, but the IQR is still much wider.

A total of 20 patients (5%) experienced severe hypoglycaemia, slightly lower than the clinical study results (7%). Simulation showed a much lower percentage of patients experiencing mild hypoglycaemia compared to clinical study results (46% vs. 74%). Hence, NS-IIT-3H succeeded in representing overall per-patient performance to what was reported clinically, although insulin administration was still much higher (99.0±66.5 U/d vs. 50.2±38.1 U/d) despite a maximum of 10 U/h limitation. These results indicate workload compliance was a major factor in the difference to clinical results, and that insulin dosing was often limited clinically compared to protocol recommendations, perhaps to levels much lower than the still relatively high 10 U/h used here.

5.4.4. Results Summary

Overall, STAR provided tighter control to an intermediate BG target, with less variability and higher safety (Figure 5.1 and Figure 5.2). This outcome was achieved for a higher number of patients. The comparison between simulation and clinical reported results of the NS-IIT has indicated very poor compliance to the protocol as defined, reflected in the average number of measurements per day (9.4 vs. 25.0 measurements per day) and the high difference in average daily insulin administration (50.2±38.1 vs. 154.0±209.2 U/d), potentially increasing hypoglycaemic episodes. These results suggest a potentially significant bias in clinical outcome for the trial due to protocol non-compliance and (also) a less effective protocol leading to poor GC for many patients, rather than the impact of GC on patient outcome. These outcomes directly contradict the interpretation of this trial in the field, noting compliance was neither recorded nor analysed in the NICE-SUGAR study [32, 46].

5.5. Discussion

5.5.1. Interpretation

The increased risk of hypoglycaemia and mortality associated with IIT over conventional control in the international multi-centre NICE-SUGAR RCT [32, 46] has been used as a gold standard to make the case against GC to lower bands, despite significant contrary results and information [135]. However, the results and methodology are widely debated [65, 121], and the lack of key comparable BG metrics is a further hindrance [91, 99]. The results presented here, using validated virtual trial simulations of the protocol on highly validated virtual patients, highlight important disparities between the original NS-IIT protocol results and clinically reported values. Comparing the virtual trials and study clinical results indicate overall safe and effective control was not delivered to all patients, potentially affecting patient, and study, outcomes. This outcome alone may explain why the study failed to replicate previous early successes in this area [38].

Simulations of the NS-IIT protocol show good safety and performance which would likely yield good patient outcomes. However, the more clinical workload feasible NS-IIT-3H design, matching clinically reported workload, had significantly increased incidence of severe hypoglycaemia despite a 10U/h limitation of insulin infusion rate, but reduced risk of mild hypoglycaemia. The measurement frequency of the original design is clearly too high and likely not manageable by clinical staff [152, 174, 179].

Hence, the negative results of NICE-SUGAR are likely a result not of GC itself, but of poor compliance to a protocol that is neither patient-specific nor clinically feasible.

The clinically reported 9.4 average measurement per day clearly reflects poor compliance to a protocol demanding 24^{+} measurements per day per protocol, potentially leading to higher variability and risk of hypoglycaemia, as seen clinically. Furthermore, the excessive, and ineffective amount of insulin administration required by NS-IIT also likely led clinical staff to adapt treatment based on their experience and clinical judgment, as reflected in the enormous difference of mean insulin administration per day and the overall per-patient mean BG CDF (Figure 5.1). As the NICE-SUGAR study does not report compliance to protocol [32, 46], it is not possible to fully or accurately assess what actually happened clinically.

In contrast, STAR virtual trial simulation results show higher performance can be achieved with moderate insulin administration. STAR consistently achieved a median 6.0 mmol/L BG in a tighter and less variable manner (Figure 5.1 and Figure 5.2). This median BG and IQR evolution over time, reflective of overall time in target band, is clearly within the accepted glycaemic ranges [60, 77]. STAR is a personalised model-based GC protocol, and compliance is very high [87]. More importantly, virtual trial results are near identical to clinical results [95, 96, 105], indicating very high compliance and good match between virtual and clinical patient cohorts.

Poor safety from hypoglycaemia is the major outcome held against IIT and GC to lower target bands. In simulations, NS-IIT and NS-IIT-3H have higher incidence of severe and mild hypoglycaemia than STAR. Clinically, the number of severe hypoglycaemic events was higher, at 7% of patients. Additionally, the percentage of patients experiencing mild hypoglycaemia (BG within 2.3-3.9 mmol/L) was much higher clinically (74%) compared to simulations (55% for NS-IIT and 46% for NS-IIT-3H), while STAR once again showed the best results with only 34%. If, indeed, NS-IIT increased the risk of hypoglycaemia compared to the conventional group, it could have been avoided by using a safer, more patient-specific protocol design. In total, clinical reported results show an important and very large 81% of patients who had a minimum BG below the mild hypoglycaemic upper threshold (BG < 4.0 mmol/L). Virtual trial simulation results on the 442 virtual patients showed 57% of them would have experienced this level of hypoglycaemia for NS-IIT, 51% for the NS-IIT-3H, and only 35% for STAR. Overall, it suggests NS-IIT

may also not be safe, in which case, despite perfect compliance, it too would likely have failed to improve patient outcomes.

The safety issue of the NS-IIT and NS-IIT-3H comes primarily from poor protocol design. While lacking patient-specificity, the original NICE-SUGAR algorithm does also not specify any insulin limits (to the knowledge of the authors), a critical safety concern. As a result, the NS-IIT insulin mean rate and SD are very large, especially for very insulin resistant patients for whom insulin effect quickly saturates and has limited to no impact, while the protocol recommends further increases in insulin doses. This issue, in turn, leads to control variability when patient state changes or patients become more insulin sensitive. However, this issue is somewhat addressed in NS-IIT-3H analysis with a specified 10 U/h maximum insulin rate limit to better match realistic control, where the mean (SD) insulin dose administered is already more pragmatic. These patients, the hardest patients to control, typically reflect the problem with highly resistant patients in ICU, where modulating (decreasing) caloric intake temporarily, as is done in STAR, effectively and safely reduces glycaemic levels in the target band, as shown in Table 5.2. This approach is also validated in other studies on STAR [87, 134].

While hypoglycaemia is the main safety concern, severe hyperglycaemia is also associated with worse outcome and higher metabolic inflammation [180]. The 5-95th percentile range of per-patient mean BG suggests 5% of patients had a mean BG higher than 9.7 mmol/L for NS-IIT (9.6 mmol/L for NS-IIT-3H) compared to 8.1 mmol/L for STAR (Figure 5.1). Hence, 5% of patients had a mean BG above the severe hyperglycaemic limit for NS-IIT, despite targeting a much lower band, a clear control failure.

Overall, this comparison of clinically reported values of the NICE-SUGAR study alongside validated virtual trial simulation results of the protocol, reflective of perfect protocol compliance with zero error, showed NICE-SUGAR failed to achieve the necessary safe and effective control for all patients in the IIT arm. Compliance to this protocol was very poor based on these results. Thus, the poor performance observed and its resultant reduced patient outcomes are most likely due to control protocol design and poor compliance.

5.5.2. Limitations

Many assumptions had to be made to simulate the NS-IIT protocols due to lack of complete protocol and/or nutrition delivery, as well as incomplete data reporting in the study. Nutrition type and total CHO intake are missing in the original NICE-SUGAR study and were likely variable across centres and/or patients. The daily clinical nutrition delivery in the simulations were based on the reported daily enteral, parenteral, and maintenance rates indicated. Content was assumed to be low CHO, a conservative choice in this comparison, and its administration was set constant daily, representing a conservative approach to modelling the unreported caloric intake, making control likely easier and reducing patient variability in this study compared to reality. Equally, a lower CHO assumption also accounts for, or averages, clinical nutrition stoppages due to hyperglycaemia.

Additionally, simulation results reflect the ideal case of full compliance to protocol. However, the NS-IIT-3H was designed to represent a more realistic control in terms of nurse workload and allows to give results matching clinical burden, mean (SD) BG achieved, and percentage BG within the different glycaemic bands. Finally, the patients used in simulation are drawn from a different, but similar, ICU cohort than those in the NICE-SUGAR study. These patients could thus be different metabolically, which may affect the insulin recommendations in simulation. However, previous work has shown similar variability in underlying SI between different ICU cohorts, showing such simulation on virtual patients to give consistent BG outcomes across very different cohorts [128].

A further limitation is the lack of reported BG metrics in the original NICE-SUGAR study around each arm. The given estimated per-patient BG distribution and cohort mean (SD) BG values reported can give indications of the differences in BG between the different groups. However, they may not reflect the underlying protocol performance in that this choice and assumed normal distribution likely tightens BG distributions and does not reflect incidence of extreme BG values. It is also influenced by the length of stay (LOS) of the patients. Additionally, reporting insulin rates using mean (SD), based on what is reported in NICE-SUGAR, is not intuitive as it does not provide any precise details on insulin rates distribution. Thus, the large corresponding SD values in Table 5.2 may seem unreasonable, but typically reflect skewed (not normal) distributions [110], which would have been better characterised using the

median and IQR values. Hence, better metrics are needed and have been called for in working group statements [90, 91, 99, 118].

Reported BG metrics from the simulations are linearly resampled hourly to allow better representation of reality, but also fair comparison between protocols. In the original NICE-SUGAR study, time-weighted BG levels are reported, which have been used in Table 5.2. Hence, the comparison between clinical BG outcome from the NICE-SUGAR study and simulations is robust and the measurement frequency have minimal impact on the results here.

5.5.3. Discussion Summary

Many reasons could thus explain why NICE-SUGAR failed. First, NICE-SUGAR is a sliding-scale controller lacking patient-specificity, while the need of patient-specific solution accounting for inter- and intra- patient variability have been widely shown [85], even before the start of the NICE-SUGAR study [162]. The poor compliance to this protocol, which could potentially have been expected, also has major impact on the results, and reflects the impact of the large potential variability arising from multi-centre RCT designs with variable results across centres [43, 47, 89]. While statistical power is often improved by such trials [144, 145], it all relies on effective and generalizable protocol design, which was not the case here. Overall, while the validity of RCT outcome on the impact of different treatments on a specific measured key outcome, such as mortality, can be erroneous if the treatment itself failed to achieve what it was designed for, and consequently induces interfering external factors, such as hypoglycaemia. In addition, metabolic control, rather than insulin itself, has been related to beneficial effects [181], whereas the poor control provided by NICE-SUGAR may have had negative effect on outcome in the intensive arm. Altogether, these issues suggest the intensive arm glycaemic and clinical outcome of the NICE-SUGAR study could have been biased simply by a poor GC protocol design, and thus the comparison with the conventional arm and the main conclusion are not valid.

5.6. Summary

Virtual trials results of the NICE-SUGAR and STAR protocols were compared to the reported NICE-SUGAR study clinical values. Per-protocol, STAR showed safer and tighter control for nearly all patients. Compared to published results, virtual trial simulation results of the NICE-SUGAR protocol showed big

differences suggesting a clinical lack in compliance to protocol. This observed poor compliance could be the reason why the intensive GC IIT arm in this major study resulted in the reported increased hypoglycaemia and increased mortality. Thus, this study suggests a different interpretation of the NICE-SUGAR results, where poor GC protocol design and added non-compliance should be accused for increased hypoglycaemia and mortality, rather than IIT and GC itself.

GC protocols need to be both safe and effective for all patients before potential clinical benefits can be assessed. Poor control can result in higher hypoglycaemia, hyperglycaemia, and glycaemic variability, all associated with worse outcomes, increased LOS, workload, and mortality, as well as cost. The need for computerised, patient-specific solutions accounting for patient variability, and achieving high quality control such as STAR is thus of paramount importance in the field of GC and offers a rule for control systems engineering.

While this study highlights the importance of the quality of GC protocols before assessing clinical outcome, it does not give a response on what target band should be used for GC in ICUs. However, it suggests that future studies, as local clinical implementation or large RCTs, aiming to respond to this question should ensure both safety and efficacy of GC protocol design before it can determine whether lower or higher BG levels are associated with beneficial clinical outcomes. This analysis thus emphasises protocol design as a key to provide safe, effective control for nearly all patients, but is it really possible to do so for all patients?

Chapter 6: Safe and Effective GC for All

Protocol design is essential to provide high quality GC outcomes. Key factors, such as patient-specificity and high compliance to protocol, are often underestimated, resulting in poor control. While this issue is important, one could wonder if the poor control provided to some patients was arose from severity of their condition and resulting prognosis for survival, rather than poor control. Specifically, perhaps those patients who die were just more complex and thus harder to control, thus linking glycaemic level and variability to outcome.

This chapter aims to determine whether it is possible to provide equal control to all patients, regardless of their clinical severity and outcome (whether they survived or not). This goal is addressed by statistically comparing the level of difficulty to provide GC between survivors and non-survivors. More specifically, SI levels and variability are analysed, as they are markers of the difficulty of control. If both groups are equally controllable, they should thus receive equal quality of control, which in turn would indicate severity should not play a role in level of GC provided.

This chapter presents results published in [135].

6.1. Introduction

The strong associations of BG level and/or variability with mortality [10, 54, 55, 57-59, 182, 183] have been used to make a case for GC. The association of moderate or severe hypoglycaemia with increased mortality [10, 11, 32, 163] similarly indicates improved control must be achieved safely, despite high inter- and intra- patient variability [58, 59, 82-85, 169].

The analysis in Chapter 5, suggested the conclusions of the NICE-SUGAR RCT could be biased. It shows poor clinical compliance to protocol may have affected the results, and thus the associated increased risk of hypoglycaemia could be a consequence of protocol design, rather than GC itself [160]. Additionally, the NICE-SUGAR protocol's lack of patient-specificity and inability to safely manage patient variability could also affect control performance and safety, where all these factors have been widely shown to be critical for success [10, 54, 55, 58, 85, 91, 99, 104, 183].

Additionally, the association of high times in intermediate bands with reduced mortality [71, 76-80] suggests control quality must be consistent over time and most (or all) patients, which only a few studies considering outcome achieved [35, 36, 45, 81]. This overall case states outcomes are largely driven by the quality and consistency of GC.

However, association is not causality. Another, equal, interpretation of these associations is that non-survivors are harder to control, and thus they have the higher glycaemic levels and variability associated with mortality. Similarly, it may be patients who die are more variable and thus more likely, under insulin control, to experience moderate or severe hypoglycaemia as a result of their underlying metabolic variability. Such patients would also have less time in intermediate bands. The equivalent case states that survivors are less variable, and thus easier to control, resulting in the more normal, consistent glycaemia associated with improved outcomes. This overall case suggests glycaemia and outcomes are driven by patient condition irrespective of GC protocol, or even that ineffective GC causes harm [62].

Separating these two interpretations would clarify the debate, research and practice around GC. In the first case, do we need better control, including any new sensors and devices, to achieve safe, effective and consistent GC for all patients in any unit? Or, in the second case, are GC and its outcomes merely a reflection of underlying patient state, and thus perhaps less necessary to control beyond a modest

lowering? In summary, is patient glycaemia and outcome (predominantly) a function of the GC achieved or driven by patient condition?

6.2. Aim and Research Question

This analysis aims to separate these two interpretations by asking the question: are patients who die harder (metabolically) to control than patients who live? If they are harder to control, then it could be considered that patient condition drives glycaemia and outcome. If not, then the quality of control could have the greater influence.

This question is addressed through a retrospective analysis of clinical data and metabolic level and variability using STAR [93, 94, 96, 123]. Lower metabolic level, captured as lower SI, indicates increased insulin is required to lower BG, which increases hypoglycaemic risk if there is variability. Greater metabolic variability, captured as greater hour-to-hour variation in SI, translates to greater outcome glycaemic variability in response to insulin. Thus, both measures capture the level of difficulty in GC, where a constant level of SI could be readily titrated to an optimal insulin dose, but unpredictable patient variability can result in excessive hyper- and hypo- glycaemia and glycaemic variability.

In short, do non-survivors have lower SI and/or greater hour-to-hour variability in SI, indicative of being harder to control compared to survivors? A positive answer would indicate the well-known associations between glycaemia and outcome are driven more predominantly by patient condition. If non-survivors were similarly difficult or easier to control than survivors, it would indicate the quality of GC achieved predominates in determining glycaemia and outcome.

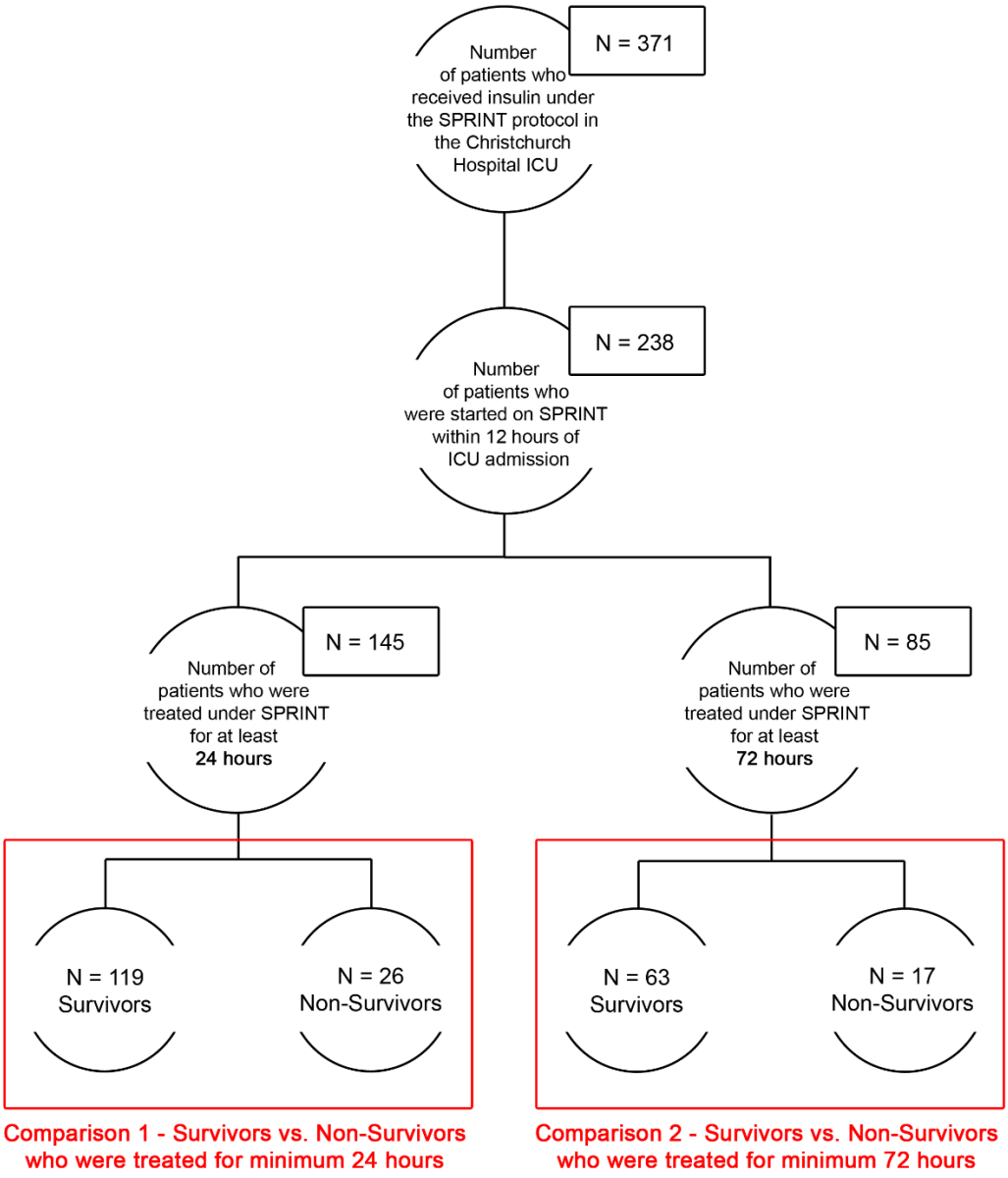

Figure 6.1 – Cohort selection from original 371 SPRINT patients. The first comparison compares survivors and non-survivors from Cohort 1, using as much data as possible, excluding very short stay patients. The second comparison uses Cohort 2 to assess the impact of competing risk due to patient dropout.

6.3. Methods

To answer the research question, metabolic state and variability is analysed using model-based SI. Key outcomes include:

- Difference and/or equivalence of SI in survivors and non-survivors
- Difference and/or equivalence of SI variability (%ΔSI) in survivors and non-survivors

Previous retrospective analysis showed SI of critically ill patients is lower and more variable during the first 24 hours of ICU stay, where SI was analysed in 6 hour blocks [82, 184]. However, differences between survivors and non-survivors or other clinical outcomes were not analysed. These outcomes are thus compared in 6-hour blocks across the first 72 hours of patient GC in the ICU.

6.3.1. Patient cohort

Retrospective clinical data from 292 patients from the SPRINT cohort presented in Chapter 4 is analysed. Figure 6.1 shows the inclusion criteria for study analysis. Of 443 GC episodes, 231 started within 12 hours of ICU admission and 145 underwent at least 24 hours of insulin therapy. This ensure there are only small differences between ICU admission and time on the SPRINT protocol and avoids any bias due to different time since ICU admission, given the evolution seen in [82, 83, 184] for the cohort as a whole. These patient episodes make up Cohort 1, with 119 (82%) survivors and 26 (18%) non-survivors. Demographics are shown in Table 6.1.

Glycaemically, survivors and non-survivors have similar times in band. Cohort median BG is statistically different (5.8 vs 5.5 mmol/L, $p<0.01$), but this difference is within clinical equivalence (explained in Section 2.4) and thus considered not clinically significant. Maximum Sequential Organ Failure Assessment (SOFA) scores on Day 1, excluding Glasgow Coma Score [38], are higher for non-survivors, as expected, and detailed breakdowns for specific co-morbidities show similar trends. All other demographics are similar except for an expected difference in Apache II score.

To assess any impact of patient dropout, Cohort 2 considers only patients who underwent at least 72 hours of GC (80 patient episodes). The first cohort assesses as much data as possible, excluding very short patients, while the second cohort assesses the impact of competing risk in the analysis of SI and mortality outcome due to patient dropout. Demographics are shown in Table 6.2, totalling 63 (79%) survivors and 17 (21%) non-survivors, and are similar to those of Cohort 1.

6.3.2. Model-based Insulin Sensitivity

Using the ICING physiological model presented in Chapter 3, integral-based fitting [131] is used to determine SI hourly from clinical BG, insulin and nutrition data. As a reminder, SI is a time-varying, treatment independent parameter characterising patient-specific metabolic response to insulin and

glucose [128]. Hence, it also reflects patient-specific general metabolic state. Consistent low SI (high insulin resistance) suggests significant stress and inflammatory state, which alleviates as the initial insult subsides [5, 7, 54, 82, 83].

SI level is determined hourly for each patient, and the forward SI variability (%ΔSI) is defined as the hour-to-hour percentage change in SI:

$$\%\Delta SI_n = 100 \times \frac{SI_{n+1} - SI_n}{SI_n} \qquad (6.1)$$

Table 6.1- Baseline data from Cohort 1 (145 SPRINT patients).

	Cohort 1	Survivors	Non-Survivors	P
N	145	119 (82%)	26 (18%)	/
Age (Yr)	67 [57, 75]	66 [57, 74]	73 [59, 78]	0.15
Gender (M/F)	91/54	75/44	16/10	1.00
APACHE II Score	20 [17, 26]	19 [16, 25]	22 [19, 31]	<0.01
First day SOFA score **Cardiac** **Pulmonary** **Hepatic** **Renal** **Coagulation**	6 [4, 8] 3 [1, 4] 3 [2, 4] 0 [0, 0] 0 [0, 0] 0 [0, 1]	6 [4, 8] 3 [1, 4] 3 [2, 3] 0 [0, 0] 0 [0, 0] 0 [0, 1]	8 [6, 8] 4 [1, 4] 3 [2, 4] 0 [0, 1] 0 [0, 0] 0 [0, 0]	0.02
ICU LOS (hrs)	113 [65, 212]	127 [65, 256]	108 [65, 154]	0.49
SPRINT Duration (hrs)	83 [44, 159]	81 [42, 168]	101.5 [55, 126]	0.93
Diabetes T1/T2 (%total)	9 / 24 (33)	8 / 21 (29)	1 / 3 (4)	1.00
Cohort BG mmol/L	5.7 [4.9, 6.7]	5.8 [5.0, 6.8]	5.5 [4.8, 6.4]	<0.01*****
Per patient BG mmol/L	5.7 [5.2, 6.2]	5.8 [5.2, 6.2]	5.3 [5.1, 5.9]	0.03
Per patient % BG in 4.4-8 mmol/L (% all BG)	82.8 [71.9, 89.5] (79.3)	82.1 [72.2, 89.3] (79.1)	83.3 [70.4, 94.4] (80.0)	0.71
Per patient % BG in <4 mmol/L (% all BG)	1.4 [0.0, 5.6] (3.4)	1.4 [0.0, 4.2] (3.0)	1.9 [0.0, 8.5] (5.0)	0.19
Num. patients BG < 2.2 mmol/L	0	0	0	/
BG measurements / day	15.8 [14.5, 17.7]	15.8 [14.4, 18.0]	15.7 [14.8, 16.2]	0.80
Per patient Median Insulin (U/hr)	3 [2, 3]	3 [2, 3]	3 [2, 3]	0.34
Per patient Median feed (g/hr)	3.2 [1.9, 4.8]	3.3 [1.9, 4.5]	3.1 [2.0, 5.3]	0.58

* indicates equivalence, as explained in Section 2.4. Data is given as median [IQR] unless otherwise indicated. P-values were computed using Fisher exact and Wilcoxon rank-sum tests where appropriate.

Table 6.2 – Baseline data from Cohort 2 (80 SPRINT patients).

	Cohort 2	Survivors	Non-Survivors	P
N	80	63	17	/
Age (Yr)	66 [54, 75]	65 [49, 74]	73 [57, 76]	0.50
Gender (M/F)	51 / 29	41 / 22	10 / 7	0.78
APACHE II Score	21 [17, 27]	21 [16, 27]	21 [17, 28]	0.60
First day SOFA score **Cardiac** **Pulmonary** **Hepatic** **Renal** **Coagulation**	7 [4, 8] 3 [1, 4] 3 [2, 4] 0 [0, 0] 0 [0, 0] 0 [0, 1]	6 [4, 8] 3 [1, 4] 3 [2, 4] 0 [0, 0] 0 [0, 0] 0 [0, 1]	8 [6, 8] 4 [2, 4] 4 [2, 4] 0 [0, 1] 0 [0, 0] 0 [0, 1]	0.11
ICU LOS (hrs)	180 [136, 371]	214 [142, 405]	142 [108, 159]	<0.01
SPRINT Duration (hrs)	155 [109, 301]	161 [126, 332]	110 [102, 151]	0.01
Diabetes T1/T2 (%total)	5 / 10 (15)	4 / 8 (12)	1 / 2 (3)	1.00
Cohort BG mmol/L	5.7 [5.0, 6.7]	5.8 [5.1, 6.8]	5.6 [4.9, 6.5]	<0.01*
Per patient BG mmol/L	5.8 [5.3, 6.2]	5.9 [5.4, 6.2]	5.4 [5.2, 6.0]	0.11
Per patient % BG in 4.4-8 mmol/L (% all BG)	84.7 [73.6, 91.7] (81.3)	84.7 [74.0, 91.7] (81.5)	83.3 [71.5, 94.8] (80.7)	0.98
Per patient % BG in <4 mmol/L (% all BG)	1.4 [0.0, 2.8] (2.5)	1.4 [0.0, 2.8] (2.0)	1.4 [0.0, 5.6] (4.2)	0.31
Num. patients BG < 2.2 mmol/L	0	0	0	/
BG measurements / day	15.1 [13.8, 16.3]	15.1 [13.4, 16.7]	15.4 [14.7, 15.9]	0.66
Per patient Median Insulin (U/hr)	3 [2, 3]	3 [3, 3]	3 [2, 3]	0.40
Per patient Median feed (g/hr)	3.3 [1.9, 4.8]	3.5 [1.9, 4.6]	2.8 [2.1, 5.6]	0.80

** indicates equivalence, as explained in Section 6.3.3.2. Data is given as median [IQR] unless otherwise indicated. P-values were computed using the Fisher exact and Wilcoxon rank-sum tests where appropriate.*

While model-based SI is used to determine whether more or less insulin should be used to lower BG levels to a safe target range, its hour-to-hour percentage change (%ΔSI) is used to assess potential risks of metabolic variability within a 1-3hourly timeframe [95, 124, 125, 185]. This variability is what makes GC difficult to achieve safely [85]. For example, at a given insulin infusion rate, a sudden increase in SI could lead to unintended hypoglycaemia, and vice-versa. It is extremely important for a GC design to assess both inter- and intra- patient variability [85]. Hence, a difference in SI levels impacts control difficulty, and also shows a difference in metabolic response to injury.

6.3.3. Analysis and Statistics

This analysis compares the evolution of SI and %ΔSI in 6 hour blocks. The CDFs for each metric are created for survivors and non-survivors over each 6-hour block. These CDFs show the overall distribution, and are exactly defined as the integral of the probability density function capturing the

histogram of the data. Therefore, they clearly define the median and any percentile likelihood (y-axis) for any given SI or %ΔSI values (x-axis).

6.3.3.1. Hypothesis Testing

Hypothesis testing is used to examine differences, with $p \leq 0.05$ used as a threshold for statistical significance. The Kolmogorov-Smirnov test is used to identify bias and shape difference in distributions of %ΔSI (Chapter 2). Although it is not certain if each family of comparisons is strictly independent (each 6 hour block may depend on surrounding blocks), for completeness and to be conservative, a Bonferroni correction for multiple comparisons is used to generalise the results (Chapter 2). In both Cohorts 1 and 2, there are 12 comparisons made bringing the significance level to $p = 0.004$ (= 0.05/12) [110].

Due to relatively large number of data points, bootstrapping, explained in Chapter 2, was used to examine the difference between median SI and median %ΔSI between survivor and non-survivor cohorts [110, 111]. For each 6-hour block, data is bootstrapped 1000 times with replacement to create resampled cohorts of similar size to the original male and female cohort sizes. The 95% CI of the difference in median SI and median %ΔSI can thus be determined. If this 95% CI does not cross zero, this difference can be considered statistically different ($p \leq 0.05$) [110]. Where this CI does not cross zero, differences in medians are statistically significant with $p \leq 0.05$ [110]. A 99.6% CI, consistent with using $p=0.004$, is considered when considering Bonferroni correction for multiple comparisons.

6.3.3.2. Equivalence Testing

Equivalence testing, developed in Chapter 2, is used to assess the impact of these differences on clinical decision making, irrespective of the underlying statistical significance (p-value) [115]. An analysis was done to determine an equivalence interval for changes in SI, as reflected by clinical significance. This interval thus defines the range within which a difference of medians cannot be distinguished due the either measurement error and/or clinical significance. Clinical significance was defined as the change in SI required to exceed BG measurement error (SD ±9.4% [186]), or to cause a change in model-based insulin dose recommendations. These calculations can be found in Appendix II. In this case, the equivalence range due to measurement error was the narrowest across the range of clinical inputs

observed. This choice provides the narrowest range and thus the most conservative or stringent test of equivalence.

The resulting equivalence range for %ΔSI is typically ≈12-15%, but is dependent on BG. Thus, any change in SI or %ΔSI within these ranges cannot be detected as different from a change due to measurement error, and are thus equivalent. Equivalence testing is independent of p-values and hypothesis testing.

Equivalence is tested for SI and %ΔSI over each 6-hour interval. For SI, the bootstrapped percentage difference in median SI is compared to the equivalence range (Chapter 2). If the 95% CI for the bootstrapped percentage difference in SI medians is within the equivalence range, then equivalence in SI is accepted (⇔). For %ΔSI, the absolute difference in median %ΔSI is examined. If the 95% CI for the bootstrapped difference in median %ΔSI is within the equivalence range, then equivalence in %ΔSI is accepted (⇔). Conversely, in both cases, if the 95% CI is outside the equivalence range, equivalence is thus rejected (×). Finally, equivalence is tested for BG in Cohort 1 and Cohort 2 as a whole, using the reported equivalence range of ±9.4%, which is one SD of the relevant BG measurement error [186]. Equivalence testing in this last case determines whether the significant differences in median cohort BG in Table 6.1 and Table 6.2 are clinically significant or not.

6.4. Results

6.4.1. Insulin Sensitivity Levels

Table 6.3 shows median SI and IQR for survivors and non-survivors in both Cohort 1 and Cohort 2 over the first 72 hours. The CDFs for SI over each 6 hour block for Cohort 1 are shown in Figure 6.2. Overall, SI level increases over time, matching [82], where non-survivors have higher SI than survivors.

In Cohort 1, the difference between median SI levels is not statistically significant (95% CI crosses zero) for the first 48 hours, except for 6-11 and 30-35 hours. By day 3, the differences become significant, except for the 66-71 hour block. With the Bonferroni correction applied, only the 6-11 and 48-53 hour blocks remain statistically different. In every 6-hour block, non-survivors have higher SI than survivors. Figure 6.3 shows results of the equivalence test for each 6 hour block. At no time do the median and

95% CI for the percentage difference of SI medians in survivors and non-survivors fall within the equivalence range. Therefore, the median SI level is never equivalent in survivors and non-survivors, regardless of p-values assessing difference.

Results are similar for Cohort 2. However, median SI is only statistically different only for hours 48-53 after a Bonferroni correction. Survivors and non-survivors are never equivalent, and SI is always higher for non-survivors in Cohort 2, who all have LOS of 3 days or larger.

Table 6.3 – SI level (L/mU/min) median [IQR] comparison between survivors and non-survivors using 6-hour blocks.

Hours		Cohort 1: 145 patients			
		Survivors (SI_S) L/mu/min× 10^{-4}	Non-Survivors (SI_{NS}) L/mu/min× 10^{-4}	Median SI_S-SI_{NS} [95% CI] L/mu/min× 10^{-4}	
Day 1	0-5	1.39 [0.50, 2.54]	1.64 [0.63, 2.63]	-0.25 [-0.60, 0.06]	×
	6-11	1.94 [1.11, 3.35]	2.58 [1.42, 3.97]	**-0.63 [-1.04, -0.11]*+**	×
	12-17	2.54 [1.42, 4.48]	3.39 [1.63, 4.79]	-0.79 [-1.46, 0.22]	×
	18-23	2.76 [1.57, 5.09]	3.22 [1.93, 5.16]	-0.42 [-0.93, 0.14]	×
Day 2	24-29	2.96 [1.65, 4.98]	3.30 [1.81, 4.85]	-0.30 [-0.73, 0.13]	×
	30-35	3.08 [1.83, 5.73]	4.34 [2.35, 7.21]	**-1.23 [-2.16, -0.20]***	×
	36-41	3.13 [1.81, 5.44]	3.42 [2.23, 5.36]	-0.29 [-1.01, 0.43]	×
	42-47	3.22 [1.81, 5.47]	4.43 [2.48, 6.24]	-0.25 [-0.94, 0.16]	×
Day 3	48-53	3.28 [1.95, 5.36]	4.83 [3.13, 8.63]	**-1.57 [-2.36, -0.97]*+**	×
	54-59	3.55 [2.03, 5.50]	4.65 [2.53, 7.27]	**-1.12 [-2.04, -0.40]***	×
	60-65	3.39 [2.18, 5.18]	4.19 [2.71, 6.83]	**-0.81 [-1.59, -0.01]***	×
	66-71	3.40 [2.43, 5.07]	3.86 [2.43, 8.30]	-0.47 [-1.43, 0.16]	×
Hours		**Cohort 2: 80 patients**			
		Survivors (SI_S) L/mu/min× 10^{-4}	Non-Survivors (SI_{NS}) L/mu/min× 10^{-4}	Median SI_S-SI_{NS} [95% CI] L/mu/min× 10^{-4}	
Day 1	0-5	1.39 [0.43, 2.45]	1.38 [0.30, 2.54]	-0.00 [-0.52, 0.57]	×
	6-11	1.90 [0.92, 3.66]	2.22 [1.15, 3.62]	-0.33 [-1.00, 0.02]	×
	12-17	2.36 [1.37, 4.48]	2.46 [1.46, 4.50]	-0.12 [-1.19, 0.61]	×
	18-23	2.63 [1.53, 4.47]	2.94 [1.87, 4.50]	-0.30 [-0.81, 0.14]	×
Day 2	24-29	2.95 [1.53, 4.52]	3.19 [1.65, 4.82]	-0.26 [-0.75, 0.22]	×
	30-35	3.04 [1.88, 5.07]	3.56 [2.24, 6.85]	-0.55 [-1.95, 0.12]	×
	36-41	3.06 [1.79, 4.94]	3.15 [2.14, 5.04]	-0.10 [-0.79, 0.51]	×
	42-47	3.21 [1.80, 5.23]	3.41 [2.93, 5.27]	-0.24 [-0.86, 0.22]	×
Day 3	48-53	3.31 [1.98, 5.30]	4.59 [3.03, 8.20]	**-1.26 [-1.84, -0.41]*+**	×
	54-59	3.59 [2.09, 5.50]	4.37 [2.43, 7.36]	**-0.87 [-1.81, -0.09]***	×
	60-65	3.45 [2.18, 5.24]	3.94 [2.62, 6.53]	-0.48 [-1.37, 0.25]	×
	66-71	3.41 [2.43, 5.21]	3.68 [2.42, 7.56]	-0.30 [-1.20, 0.24]	×

*Hours where the medians are statistically different (95% CI on difference in medians does not cross zero) are marked with *. Equivalence is marked with ⇔ and non-equivalence with an ×. Differences remaining significant after a Bonferroni correction are marked with + (99.6% CI on difference in medians does not cross zero).*

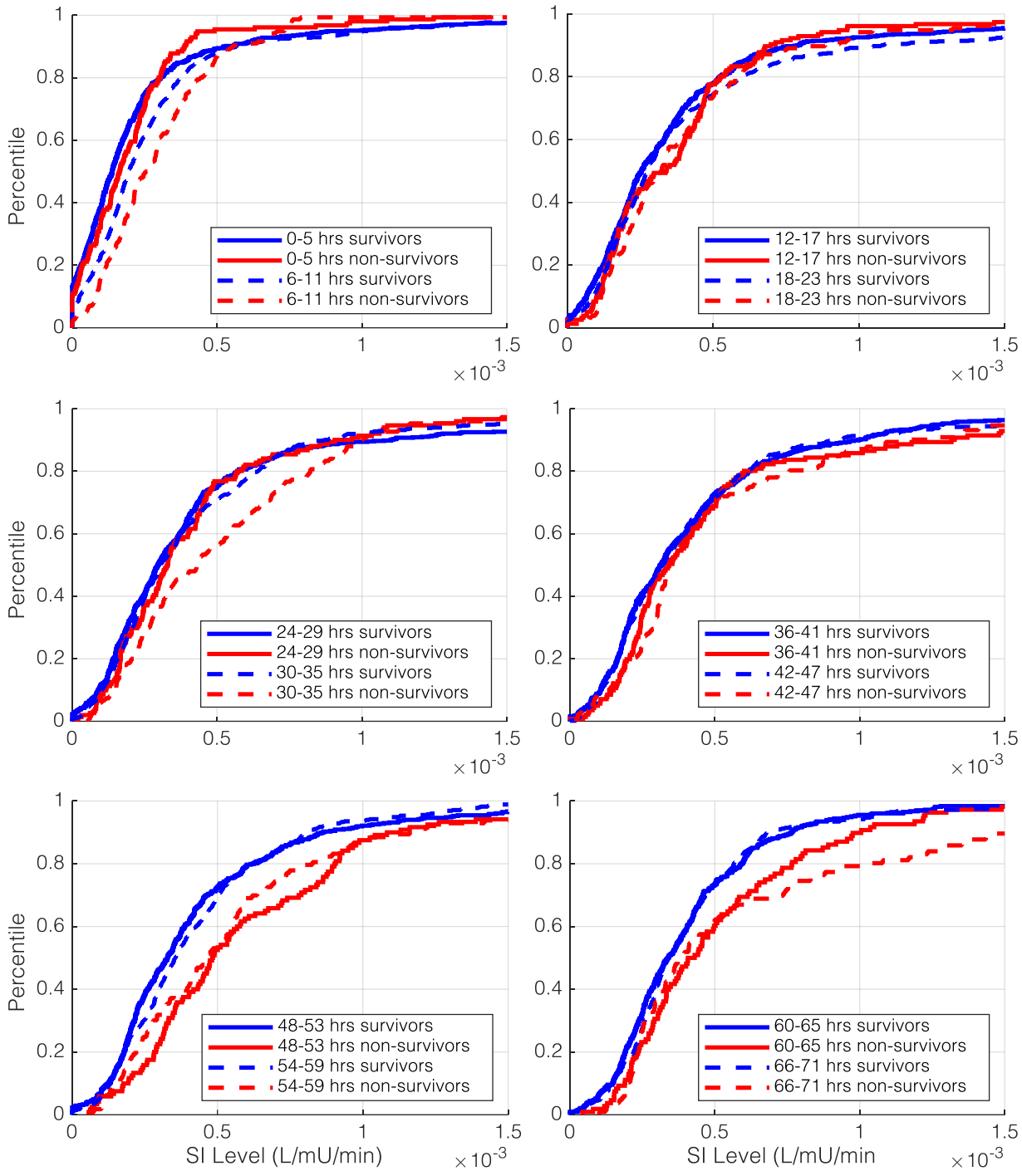

Figure 6.2 – Cohort 1 cumulative SI levels over 6-hour time intervals for the first 72 hours of glycaemic control. At any level of SI, the y-axis gives the percentage of SI values (decimal percentile) below this level.

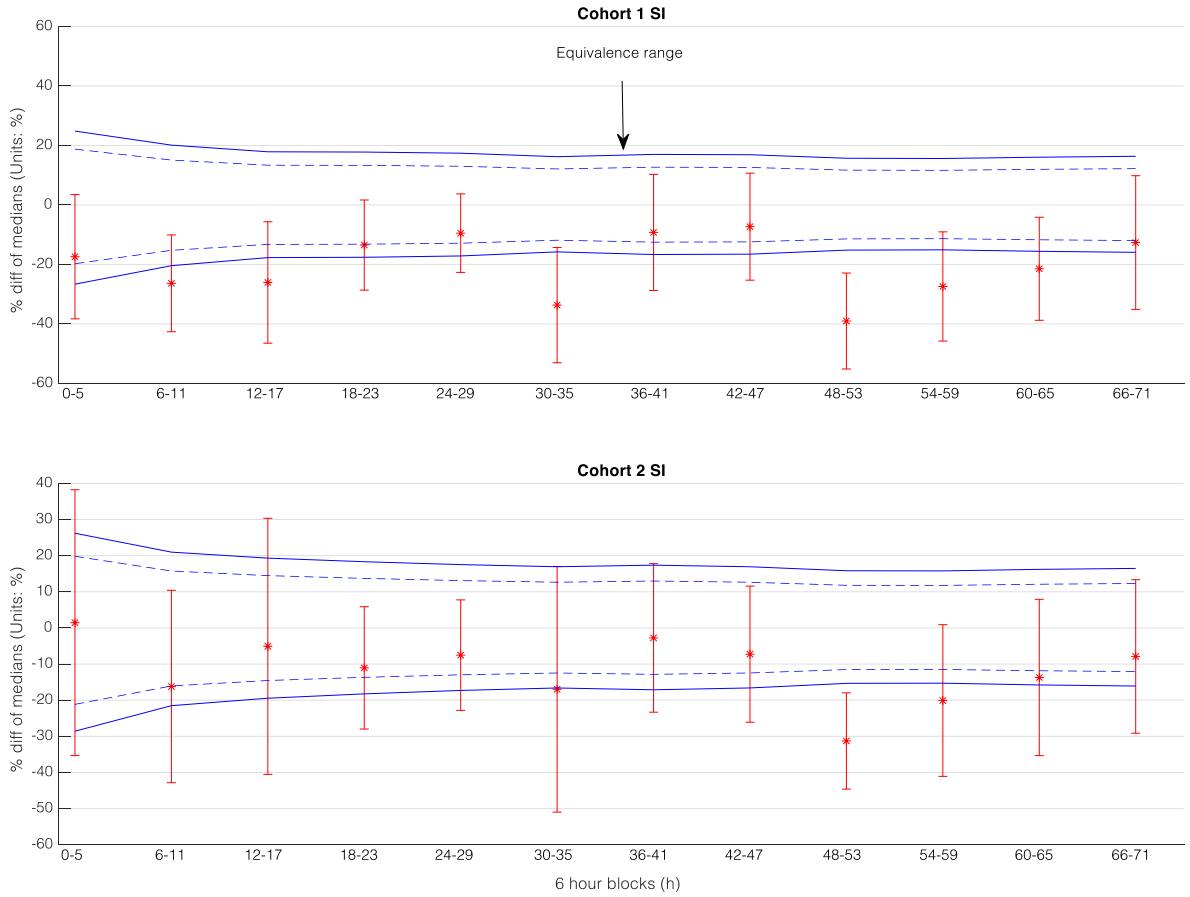

Figure 6.3 – Equivalence testing on SI for each 6 hour block for Cohort 1 and Cohort 2. The solid blue lines give equivalence ranges for 9.4% BG error [186] and the blue dotted lines a smaller 7% error reported for the device used in highly controlled tests [187]. Equivalence is accepted (⇔ in Table 6.3) if the 95% CI (bars) of bootstrapped percent difference in median SI is within the equivalance range, and rejected otherwise (×).

Figure 6.4 shows the evolution of median [IQR] SI and BG over time, between survivors and non-survivors, for Cohort 1 and Cohort 2. In both cohorts, SI is higher for non-survivors, as reflected in Table 6.3, and this difference is greater as control progresses. In terms of BG, survivors and non-survivors have similar levels for most hours. Equivalence testing on overall BG distributions between survivors and non-survivors shows the median and 95% CI of the percentage change in median BG are 5.3 [2.6, 7.1] for Cohort 1 and 3.5 [0.9, 5.3] for Cohort 2, which is well within equivalence ranges of 7.0-9.4%. Thus, while the differences are statistically different, it confirms the differences in the median BG values in Table 6.1 and Table 6.2 are not clinically significant. It is important to note these two figures do not necessarily reflect SI hour-to-hour variability at a per-patient level. Two patients could have equal variability in a 6-hour period but at different hours, and thus appear different in SI level, which explains the need of a separate %ΔSI analysis assessing the hour-to-hour variability.

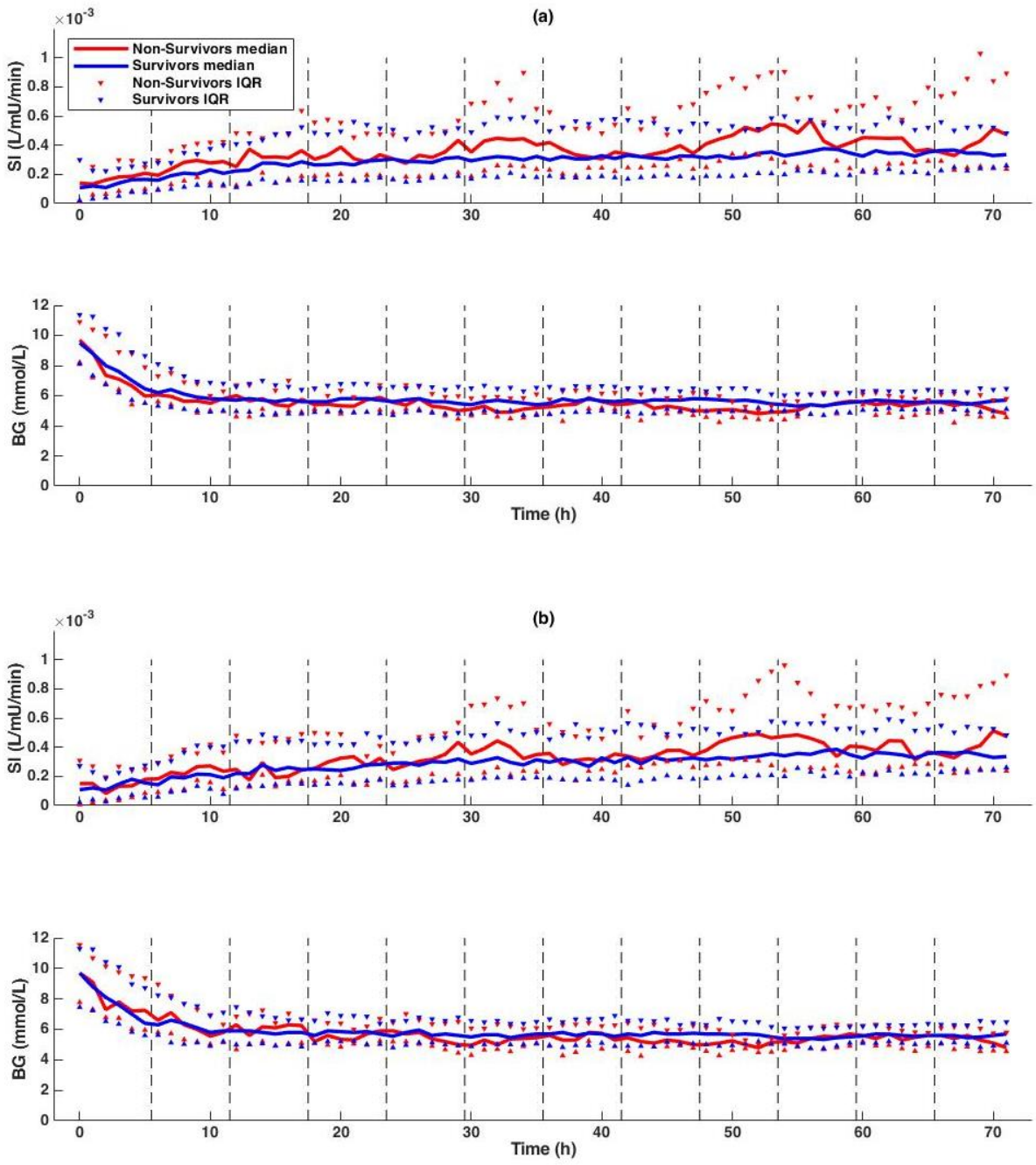

Figure 6.4 – Median [IQR] evolution of SI and BG over time for survivors (blue) and non-survivors (red) in Cohort 1 (a) and Cohort 2 (b).

6.4.2. Insulin Sensitivity Variability

Results for %ΔSI are shown in Table 6.4 and Figure 6.5. Overall, SI variability decreases over time (IQR narrows) for both survivors and non-survivors, matching [82]. In both Cohort 1 and Cohort 2, %ΔSI is not significantly different ($p \geq 0.11$ in 11/12 blocks), especially when a Bonferroni correction for multiple comparisons is made ($p < 0.004$ correction threshold). The 95% CI on median difference in %ΔSI (bias only) can only be considered significant for the 36-41 and 42-47 hour blocks in Cohort 1, and for the 18-

23 and 24-29 hour blocks in Cohort 2 (bootstrapping, last column of Table 6.4), but these significant differences do not hold when a Bonferroni correction is made (99.6% CI). In all cases, these differences were not clinically significant. As shown in Figure 6.6, the median and 95% CI change in %ΔSI difference is always within the equivalence range for both Cohort 1 and 2. Therefore, SI variability assessed as %ΔSI in survivors and non-survivors is equivalent in every 6-hour block to 72 hours.

6.4.3. Key Results

In summary, the key results are:

- SI level is not equivalent in any 6-hour block within the first 72 hours of GC, and is sometimes statistically different between survivors and non-survivors.
- SI level is higher in non-survivors than survivors in every 6-hour block for the first 72 hours, and this difference becomes statistically significant as GC progresses.
- SI variability is equivalent between survivors and non-survivors in any 6-hour block within the first 72 hours of GC.
- Patient dropout has no impact on results as Cohort 2 has the same key outcomes.
- Major results are consistent irrespective of whether a Bonferroni correction for multiple comparisons is applied.

Thus, while survivors and non-survivors differ in their absolute SI, with non-survivors having higher SI, they are equivalent in their hour to hour variability (%ΔSI).

Table 6.4 – %ΔSI (%) median [IQR] comparison between survivors and non-survivors using 6-hour blocks.

Hours		Cohort 1: 145 patients				
		Survivors (%ΔSI_S) %	Non-Survivors (%ΔSI_{NS}) %	KS-Test p-value	Median %ΔSI_S-%ΔSI_{NS} [95% CI] %	
Day 1	0-5	1.46 [-29.26, 54.74]	11.67 [-20.84, 56.41]	0.67	-8.12 [-16.22, 4.67]	⇔
	6-11	7.37 [-14.66, 42.05]	9.47 [-11.45, 27.98]	0.53	-1.31 [-6.22, 5.74]	⇔
	12-17	5.21 [-11.87, 30.89]	6.69 [-14.89, 42.15]	0.62	-0.98 [-9.46, 7.26]	⇔
	18-23	3.24 [-16.02, 26.92]	-0.63 [-12.21, 16.37]	0.12	3.72 [-1.99, 8.56]	⇔
Day 2	24-29	2.79 [-13.36, 23.35]	5.37 [-9.42, 23.52]	0.30	-2.70 [-8.60, 3.29]	⇔
	30-35	1.76 [-15.13, 23.46]	1.57 [-11.32, 24.75]	0.78	0.34 [-8.54, 6.75]	⇔
	36-41	1.92 [-12.19, 16.87]	-4.01 [-15.63, 11.26]	0.04	6.10 [0.35, 10.70]*	⇔
	42-47	-0.10 [-12.71, 17.98]	5.46 [-10.91, 21.91]	0.14	-5.66 [-11.61, -0.43]*	⇔
Day 3	48-53	1.57 [-10.74, 16.82]	3.41 [-7.30, 14.99]	0.30	-2.12 [-7.41, 1.77]	⇔
	54-59	0.67 [-11.68, 15.80]	-3.13 [-19.08, 11.65]	0.35	3.37 [-1.77, 8.20]	⇔
	60-65	2.39 [-12.39, 17.03]	4.89 [-8.88, 21.88]	0.45	-2.50 [-9.06, 3.35]	⇔
	66-71	1.26 [-9.80, 12.87]	3.78 [-8.82, 15.48]	0.35	-2.76 [-8.66, 2.80]	⇔
Hours		**Cohort 2: 80 patients**				
		Survivors (%ΔSI_S) %	Non-Survivors (%ΔSI_{NS}) %	KS-Test p-value	Median %ΔSI_S-%ΔSI_{NS} [95% CI] %	
Day 1	0-5	0 [-29.44, 43.57]	0.98 [-20.90, 57.81]	0.78	-0.98 [-16.02 5.93]	⇔
	6-11	8.80 [-14.66, 48.55]	10.59 [-17.24, 39.20]	0.90	-2.17 [-11.46 6.83]	⇔
	12-17	2.38 [-13.18, 29.19]	2.92 [-15.99, 38.92]	0.89	-0.02 [-11.00 9.19]	⇔
	18-23	4.09 [-14.80, 26.14]	-2.13 [-11.66, 15.29]	0.11	6.16 [0.10 12.10]*	⇔
Day 2	24-29	1.32 [-13.48, 20.79]	10.39 [-8.97, 25.86]	0.02	-9.23 [-14.18 -1.38]*	⇔
	30-35	0.13 [-15.56, 21.72]	3.08 [-13.33, 23.03]	0.89	-2.38 [-10.72 4.79]	⇔
	36-41	2.54 [-12.13, 18.20]	0.40 [-11.95, 15.13]	0.39	2.95 [-1.86 9.51]	⇔
	42-47	1.37 [-13.37, 22.76]	2.42 [-11.83, 14.68]	0.46	-1.02 [-7.15 5.68]	⇔
Day 3	48-53	0.88 [-10.32, 16.63]	3.16 [-7.25, 14.62]	0.30	-2.37 [-7.73 1.79]	⇔
	54-59	0.72 [-10.36, 14.28]	-1.17 [-19.08, 12.68]	0.32	2.69 [-3.18 7.71]	⇔
	60-65	2.58 [-10.54, 16.38]	4.04 [-9.03, 21.96]	0.39	-1.89 [-8.44 3.76]	⇔
	66-71	1.26 [-9.74, 11.54]	3.72 [-9.06, 14.81]	0.40	-2.62 [-8.81 3.12]	⇔

Hours where the medians are statistically different (95% CI on difference in medians does not cross zero) are marked with *. Equivalence is marked with ⇔ and non-equivalence with an ×. Differences remaining significant after a Bonferroni correction are marked with + (99.6% CI on difference in medians does not cross zero).

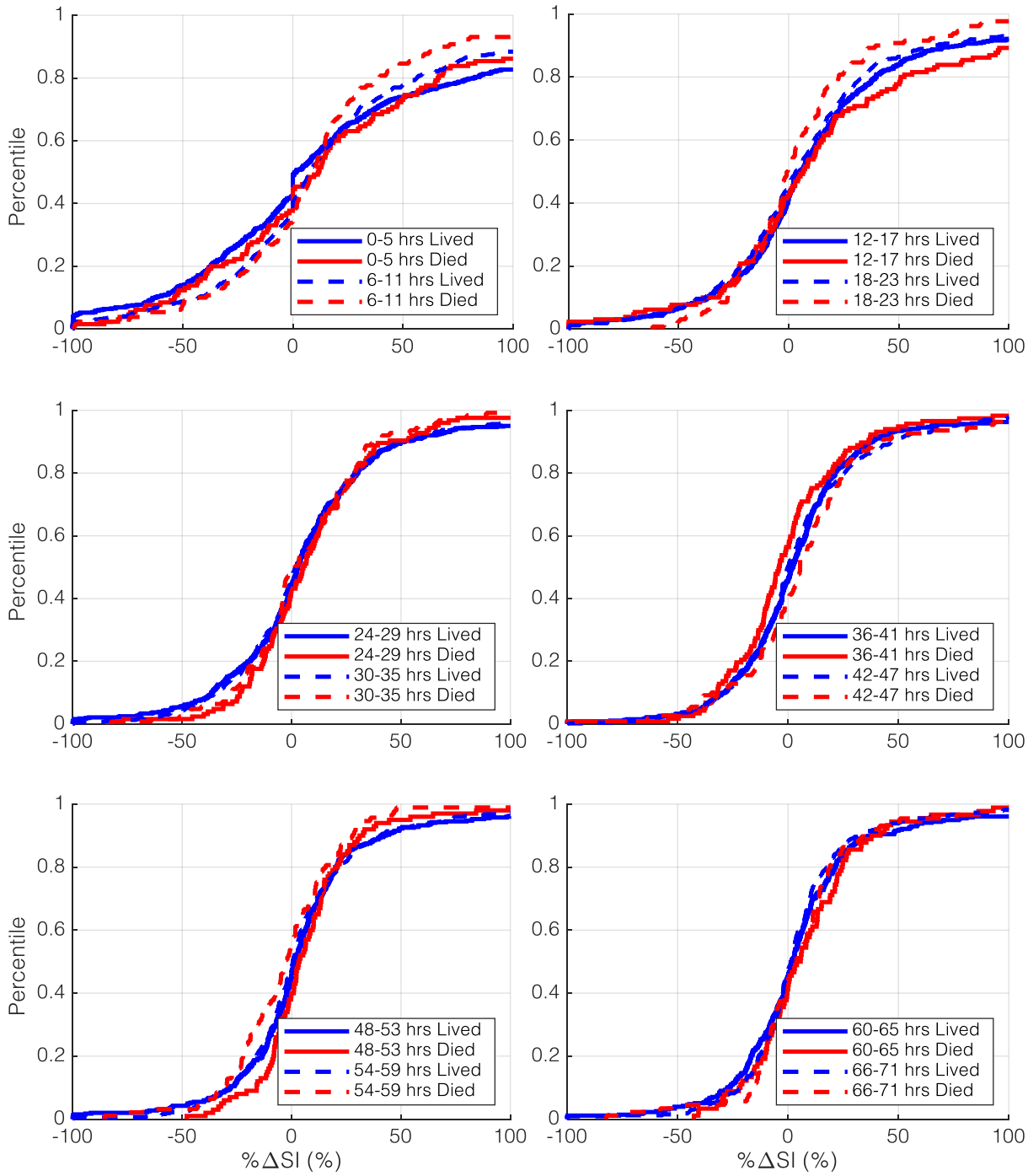

Figure 6.5 – Cohort 1 cumulative hour-to-hour percentage changes in SI level over 6 hour time intervals for the first 72 hours of glycaemic control. At any level of %ΔSI, the y-axis gives the percentage of %ΔSI values (decimal percentile) below this level.

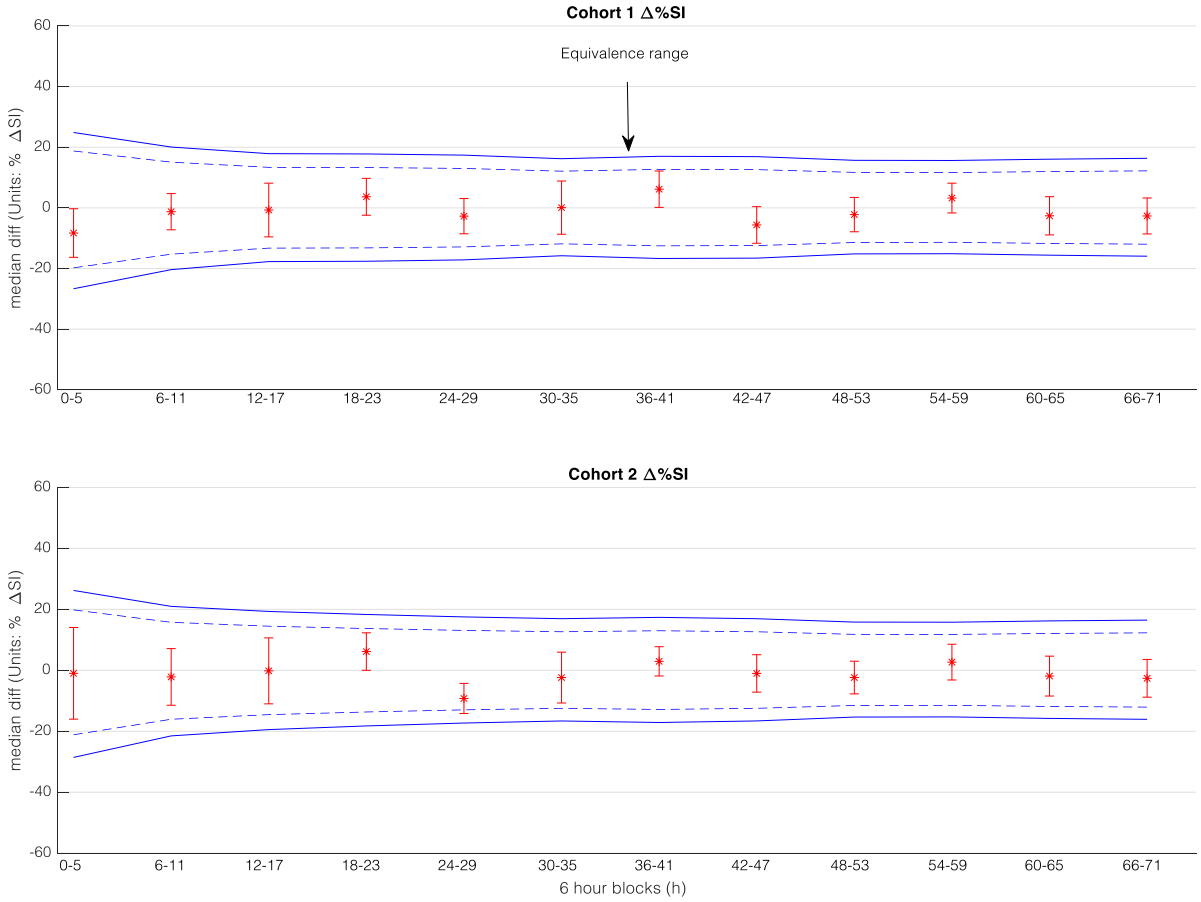

Figure 6.6 – Equivalence testing on SI variability (%ΔSI) for each 6 hour block for Cohort 1 and Cohort 2. The solid lines give equivalance ranges for 9.4% BG error [186] and the dotted lines a smaller 7% error reported for the device used in highly controlled tests [187]. Equivalence is accepted (⇔ in Table 6.4) if the 95% CI (bars) of bootstrapped difference in median %ΔSI is within the equivalance range, and rejected otherwise (×).

6.5. Discussion

6.5.1. Primary Question

Patient-specific SI and SI variability metrics are used to assess underlying controllability between survivors and non-survivors. Both statistical difference and equivalence were tested in comparing these cohorts. Statistical difference ($p < 0.05$) tests whether the data come from similar or different distributions. In contrast, equivalence tests whether these values are clinically or physiologically equivalent, regardless of p-value.

SI was statistically different between survivors and non-survivors for 5 of 12 6-hour periods. However, cohorts were never clinically equivalent in SI for any period. Non-survivors had higher SI in every period, suggesting slightly lower insulin doses would be required to achieve normo-glycaemia, which is also seen in the clinical results in Table 6.1. Key results were the same for both cohorts examined.

SI variability (%ΔSI) was equivalent between survivors and non-survivors for every period, and only statistically different in 2 periods. Equivalent variability under the same GC protocol would be reflected in similar times in glycaemic bands and in glycaemic levels for both survivors and non-survivors, as seen in Table 6.1. The results were the same for both cohorts. Median BG was higher in survivors (5.5 vs 5.8 mmol/L; $p < 0.01$ for both cohorts), but this difference is shown to be clinically equivalent in terms of measurement error and, in addition, would not change the clinical interventions.

While SI level tends to determine the total insulin dose titrated, it is variability that determines the risks of insulin therapy and overall controllability. Overall, similar to higher SI for non-survivors and equivalent variability suggest survivors and non-survivors are equally controllable given an effective GC protocol. This outcome in turn suggests the association between glycaemia and outcome is thus predominated by the quality of GC achieved, and not underlying patient variability. This result is important and has significant important clinical implications for GC study design and practice.

6.5.2. Validity of SI Metric

The results rely on the validity of the model-based estimate of SI extensively used here to compare both cohorts. The reliability of the SI metric is determined by the underlying data and the ability of the model to capture key glucose-insulin dynamics. The ICING model used here is structurally very similar to the dynamic insulin sensitivity and secretion test model, for which the SI metric has correlated well with the gold standard euglycaemic clamp SI values [136, 137, 155, 156], as have other very similar models using SI metrics and pharmacodynamics used here [140]. The ICING model and its SI metric have been successfully and safely used to guide insulin therapy across different adult [87, 95, 96, 123] and neonatal [153, 157] intensive care settings and delivery methods. These clinical results suggest the model is able to capture and account for all major glucose-insulin dynamics, making the SI parameter, and its guiding of care via forward prediction, clinically useful.

In addition, treatment independence of the SI parameter has been assessed using clinical data from independent, matched patient cohorts [94, 128]. In the first case, two cohorts and protocols (Glucontrol [47] and SPRINT [81]) from Liège, Belgium, and Christchurch, New Zealand, were simulated with both protocols, and their glycaemic level and variability compared to those obtained clinically. Consistency in simulation results across cohorts and high similarity in stochastic plots of SI variability further validate

the treatment and cohort independence of SI [86]. In the second case, this similarity and cross validation was repeated across three medical ICU cohorts, further validating these outcomes [128]. Recent work suggests it is an underlying similarity in SI variability, independent of absolute SI level, driving GC outcomes [86, 94, 128]. This similarity thus also drove the observed consistency between clinical results using this model and SI metric for GC in two very different ICUs [87].

Moreover, SI has been shown to assess and reflect clinically expected changes in SI and metabolism for important intensive care interventions. The impact of glucocorticoids [26] and β-blockers [82] on SI level and %ΔSI was shown to be limited in the context of the SPRINT protocol. More specifically, insulin and nutrition inputs were not statistically different in this study between survivors and non-survivors (Table 6.1, $p>0.34$), where increasing insulin use would reflect increased insulin resistance (lower SI). These results thus suggest any glucocorticoids-mediated influence on SI does not have any net impact on the two groups, as there was such a difference in Pretty et al [26]. Additionally, the impact of exogenous nutrition and incretin effects, seen in changes in SI [188], the impact on SI from haemodialysis altering insulin clearance [189], and, finally, the insulin resistance observed on and off therapeutic hypothermia [190], were all assessed using hourly identified SI based on the same model. Each of these studies demonstrates the ability of SI and its changes to reflect clinically expected outcomes, and correlates with expectations for the given intervention.

Other factors, such as insulin administration form (bolus vs. continuous dosing), have little impact on the hourly calculated SI value. In this study, both survivors and non-survivors were treated with boluses, eliminating any potential effect for this comparison. Glucose sensor errors could have a more measurable impact on SI calculation [170], but the same glucometers were used for all patients, similarly ameliorating this affect. Continuous Glucose Monitors (CGM) have delivered observations indicating greater apparent spontaneous variability in BG levels than seen with typical intermittent sampling. However, it is important to note a major part of this CGM-observed BG variability is not due to patient metabolism but directly related to sensor drift, changes in the in-situ environment of the sensor, patient position, and other factors [191-198]. Thus, what is captured by a CGM may be either real or an artefact, or some combination. However, differentiating these systemic errors from real BG variability is not currently possible without another reference measurement at a similar rate. As a result, the hourly determined SI values used here are appropriate, particularly to the measurement rate in the data, which

cannot capture any real glycaemic variability in the data that occurs and resolves between measurements. Hence, the overall approach used here is appropriate to the data and its sampling rate, and does captures very high levels of variability, as seen in Figure 6.5 with changes in SI up to 640%. Two example of SI profiles over time, indicating the actual variability possible, are shown in Appendix I.

Glucose complexity has been associated with mortality [199, 200], but cannot be measured at the bedside in real-time like glycaemic levels, time in band, or variability. Equally, there is not the strong physiological evidence to support this association compared to the existing evidence for the other metrics considered. Finally, there are questions about its proper use in analysing continuous glucose data to create these associations [201, 202].

The presented results suggest non-survivors have higher SI, which at first appears counter-intuitive. However, it can be hypothesized some non-survivors may have had weaker inflammatory immune responses and/or weaker inflammatory counter regulatory response to insult. While literature commonly points to increased inflammatory markers in non-survivors (e.g. [203, 204]), there is evidence of instances where compromised immune response leads to increased mortality (e.g. [205-208]). These physiological responses (both inflammatory [30, 209-213] and counter-regulatory [5, 209, 214-216]) drive hyperglycaemia via the inflammatory marker induced actions that reduce the SI values analysed here. They are also two of three major drivers of hyperglycaemia, the third being high glucose itself. Hence, weakened responses in those who die would lead to slightly higher SI, and thus may be the cause of the slightly higher SI and slightly lower, clinically speaking, insulin use in this cohort. However, we do not have evidence to prove this hypothesis, but it would make a good hypothesis for a future study.

In particular, SI is ≈20% higher on average for non-survivors, ranging from ~9-40% over the 6-hour time periods, which is at or within the level of change in SI required to induce, in SPRINT, a 1U/hour change in insulin dose, considering a median of 3U/hr (See Figure A2.5 of Appendix II). Thus, this difference changes few interventions, as seen in Table 6.1 (median [IQR] of 3 [2, 3] U/hr for both survivors and non-survivors), where feed is also similar. Finally, excluding dropouts in Cohort 2, the differences remain, but are much smaller (≈12%). Thus, while SI is higher for non-survivors and not equivalent to SI of survivors, based on the most conservative estimate (percentage change in SI to reach 9.4% BG

measurement error), this difference in SI does not have a significant clinical impact in terms of interventions, where a ~20-25% change in SI is required to change an intervention (See Figure A2.5 of Appendix II).

One advantage of the model-based SI used here is it accounts for all insulin and nutrition inputs, and resulting changes in glycaemia, allowing the SI metric to reflect the underlying ability of the body to utilise insulin for glucose uptake. Using SI thus allows an objective numerical analysis to be carried out, and for results to generalise to other mixed-ICU populations.

6.5.3. Advantages and Limitations

A first potential limitation of this work is, as with all models, the ICING model has ranges for BG and nutrition-insulin interventions in which it is most accurate [217]. These ranges span what is typically observed in the Christchurch Hospital ICU, including BG within the 4 – 10 mmol/L range, and insulin and nutrition treatments within 0 – 10 U/hr and 20 – 120% of GF, respectively. If this analysis were to be repeated in ICUs or with protocols where treatments may commonly be given outside of these ranges, or where persistent hyper- and/or hypo- glycaemia were common, there would be greater potential for analytical error. However, in this case, the clinical data and inputs all fall within the ideal range for the ICING model.

Another limitation is this analysis would be difficult to repeat with data from other, larger studies, for example, because of this lack of detail and/or temporal resolution of the GC data collected [99]. Additionally, this study is limited in its retrospective nature, and that it was performed on data from a single centre. However, the data covers a relatively large generalised patient cohort spanning several years of clinical practice, and is of high quality as explained in Chapter 4. Illness and injury can affect the inflammatory response, and thus the SI. The analyses cohorts were therefore selected on the basis of starting GC within the first 12 hours of ICU stay to reduce the effect of time-varying degrees of illness and injury on the time-varying analysis of SI.

6.6. Summary

Using strong, robust statistical analysis the results of this study show equivalent metabolic variability between survivors and non-survivors and that non-survivors have higher SI. These results are based on a numerical, objective, model-based SI metric, which takes into consideration both nutrition-insulin inputs and metabolic outcomes. The underlying data cohort is derived from a mixed-medical ICU, and as previous work has shown consistency in variability across different cohorts, countries, and centres, it is likely that the results of this study are not specific to the original data set. Given SI variability is really what makes GC hard to achieve safely, these results suggest glycaemic outcomes and differences between survivors and non-survivors are thus more a function of the control provided, rather than underlying metabolic condition. This outcome has implications for future study and protocol design in this area.

Once again, metabolic variability has been shown as playing a key role in the quality of GC outcomes. This analysis already gives some insights on the differences in SI levels and SI variability that may exist between patients. Specifically, while inter-patient variability (SI levels) cannot be considered equivalent, intra-patient variability (SI variability) appears so. Whether these differences (or similarities) hold across any sub-groups of patients could help further understand variability, and generalise these observations.

Chapter 7: Understanding Variability

GC in the ICU has been shown to be challenging due to inter- and intra- patient variability, often leading to increased risk of hypoglycaemia. Accounting for patient-specific metabolic variability is thus one of the keys to provide safe, effective control. Most importantly, hour-to-hour SI variability, reflective of intra-patient variability, makes high GC performance difficult to achieve safely. In Chapter 6, survivors and non-survivors were shown equally hard, or easy, to control, given their intra-patient variability was equivalent, indicating quality of GC was the predominant factor in glycaemic and thus clinical outcomes, rather than

This chapter aims to better understand inter- and intra- patient variability across different sub-groups of patients, and determine whether the conclusions of Chapter 6 are also true for other specific demographic characteristics. Specifically, recent work showed higher insulin resistance (lower SI) in preterm girls based on differences in insulin secretion [218-220]. This study aims to show whether a difference in inter- and intra- patient metabolic variability between sexes exists in adult ICU patients, and assesses the impact on GC and metabolic response to injury. Any significant difference would suggest GC design should consider sex differentiation to provide personalised care.

This chapter presents results published in [221].

7.1. Introduction

Patient-specific solutions using key physiological parameters to tailor control for each patient individually, including risk assessment for GC, can improve control and patient outcomes [90, 91, 98]. Such controllers exist, and have successfully shown safe, effective control while targeting lower glycaemic ranges [43, 87, 107, 108], without sacrificing nutrition delivery or other care aspects [134].

In the previous Chapter, equivalence testing on SI levels and variability was analysed between survivors and non-survivors to understand whether these subgroups are more or less difficult to control [135]. The main outcome of this analysis showed non-survivors had higher SI levels compared to survivors, and this difference was not clinically equivalent. However, SI variability between these cohorts was always clinically equivalent. These results suggest GC outcome, and thus associated mortality, is function of protocol design, rather than patient condition. Thus, high levels of safety and performance should be able to be achieved in a mixed ICU cohort, regardless of the severity of injury or eventual outcome, which is critical to seeing potential benefits [38]. These outcomes also confirm the importance for a GC design to address metabolic variability correctly, which is really what makes safe, effective GC hard to achieve [85, 135].

While quality of GC should not be influenced by patient outcome, it is possible other metabolic differences could influence control if differences in patient-specific metabolic stress response existed. In particular, a previous study on neonatal ICU patients showed greater endogenous insulin secretion in girls, suggesting a higher insulin resistance [218-220]. The results suggest a difference exists between sexes in neonates. However, no analysis, to the authors knowledge, clearly analysed any sex related differences in the context of GC in adult ICU.

Women have been clearly under-represented in clinical trials [222, 223]. In the 1980s-1990s, the lack of women included in trials was recognised [224], despite consuming 80% of pharmaceuticals in the US at that time [225, 226]. In particular, differences in how women metabolise or clear some drugs has led to significantly different and unintended concentrations, which should necessitate different dosing instructions [227]. However, their higher metabolic variability was seen as a potential outcome bias, and, in consequence, induced a male bias and preclinical and clinical research [228].

In this chapter, retrospective data are used to analyse SI levels and variability between males and females, and to understand whether there exists a difference in these subgroups. Similar to the previous chapter [135], a significant difference or equivalence could help understand whether GC is different and/or more difficult between males and females. Equally, given the impact of metabolic stress response on metabolism, it could also show whether a difference exists between the sexes in metabolic response to injury, which is currently unknown. If so, it would provide guidance on whether GC should explicitly consider sex differentiation in protocol design or via personalised care.

7.2. Methods

Patient cohort and statistical analysis are similar to those developed in Chapter 6. From the SPRINT cohort (Chapter 4), only the 145 patients (39%) who started GC within 12 hours after ICU admission and received insulin for a minimum of 24 hours are used to avoid any bias due to different time since ICU admission. This specification ensures a similar starting time and progression from insult toward recovery for all patients, and thus eliminates a potential source of bias or error in results. In these 145 patients, 91 (63%) are males and 54 (37%) are females, which is a typical breakdown in ICU cohorts. Demographic characteristics are summarised in Table 7.1.

As in the previous analysis of Chapter 6, identified SI and its hour-to-hour percentage change (%ΔSI) is calculated for each hour. Because a statistically significant difference ($p<0.05$) can have minimal impact clinically and would be too small to affect decision making, equivalence testing (Chapter 2) is used [110-112]. Equivalence testing assesses difference based on clinical significance and determines whether this difference in median SI and median %ΔSI is within a clinically set equivalence range [115].

While the raw data in the original cohort as presented are analysed first (91 males or 63% versus 54 females or 37%), the analysis was repeated using bootstrap resampled sub-cohorts of the same size. This approach allows a fair comparison to ensure no bias results from the specific patients and proportions in the original cohort [110, 111]. New resampled male and female cohorts (N=50 each) were randomly created by choosing patients from the original cohorts with replacement. In these cohorts, 8 patients (16%) were picked from patients with T2D so this factor was also balanced. This process was repeated 100 times, and hypothesis and equivalence testing on SI and %ΔSI were undertaken each time. The percent (%) of times the null hypothesis was rejected and equivalence accepted is calculated

for each 6-h block. This secondary analysis ensures no bias due to proportions or specific patient subsets, adding robustness to the overall results. Note, it does assume the patients in each group are representative of the range of behaviours, which can be further confirmed by consistency of results over the bootstrapped cases to assess any impact of outlying patients.

Table 7.1 – Demographics summary of male and female cohorts from 145 SPRINT patients.

	Males	Females	P value
# patients	91	54	
Age	67 [57 77]	67 [58 74]	0.63[a]
Mortality	18%	19%	1.0[b]
APACHE II score	20 [16 27]	19.5 [17 26]	0.98[a]
First day SOFA score	6 [4 8]	5.5 [4 8]	0.46[a]
ICU LOS (h)	108 [67.2 188.4]	127.2 [64.8 213.6]	0.91[a]
SPRINT duration (h)	83 [45.5 157.3]	86.5 [39 167]	0.81[a]
T2DM (%)	13 (14%)	11 (20%)	0.4[b]
Cohort BG (mmol/L)	5.6 [4.9 6.6]	5.9 [5.0 6.9]	<0.01[a]*
Per patient median BG (mmol/L)	5.65 [5.16 6.14]	5.99 [5.3 6.4]	0.06[a]
% BG 4.4-8.0 mmol/L	83 [72 90] (68.2)	82 [67 89] (62.7)	0.3[a]
% BG < 4.0 mmol/L	1.4 [0 5.5] (2.7)	1.4 [0 6.9] (3.1)	0.42[a]
%BG < 2.2 mmol/L	0 [0 0] (0)	0 [0 0] (0)	NA
BG measurements/day	15.8 [14.4 17.5]	15.7 [14.5 18.2	0.47[a]
Median insulin (U/h)	3 [2 3]	3 [2 3]	0.26[a]
Median feed (g/h)	3.5 [2.1 5.5]	2.8 [1.8 3.9]	<0.01[a]
Median feed(%GF)	51 [30 80]	51 [30 75]	0.61[a]
GF (g/h)	6.54 [6.54 7.41]	5.2 [5.2 5.7]	<0.01[a]

*Statistical difference is shown using ([a]) the Wilcoxon rank-sum test or ([b]) Fisher exact test where appropriate. P-values are not adjusted for multiple comparisons. Per-patient median [IQR] is given where appropriate. T2DM = Pre-diagnosed type 2 diabetes, GF = goal feed, and BG = blood glucose. * indicates clinical equivalence regardless of statistical significance.*

7.3. Results

Overall SI CDFs for males and females are shown in Figure 7.1. Clearly, the female cohort is more resistant than men (lower SI levels). SI level comparison results between males and females for every 6-h block are detailed in Table 7.2 and shown in Figure 7.2. SI levels increase over time in both cohorts, as expected [82, 83, 135]. The 95% CI of difference in median levels between male and female never crosses zero, suggesting the difference is statistically significant not only overall, but also for each 6-h block. Considering the Bonferroni correction, 60% (7/12) of the 6-h blocks remain significantly different.

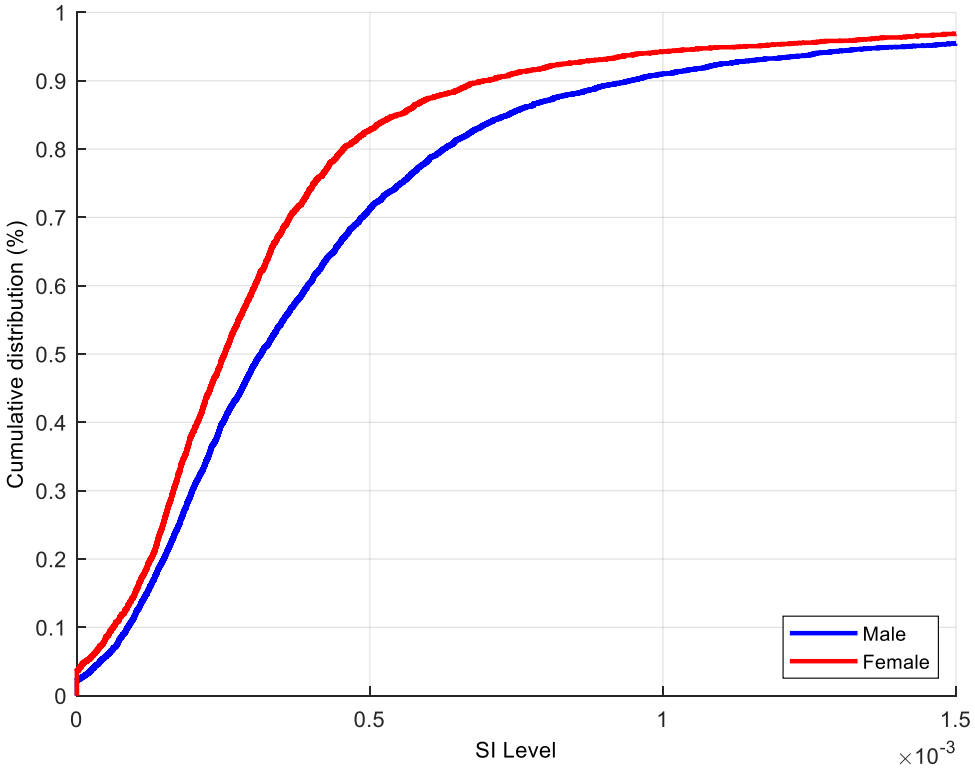

Figure 7.1 – Overall cumulative SI levels (L/mU/min) between male and female cohorts.

Table 7.2 – Median [IQR] SI levels comparison for the first 72 hours between male and female cohorts using 6-hour blocks.

Hours	Male Cohort SI (×e-4)	Female Cohort SI (×e-4)	Median SI_M-SI_F [95%CI] (×e-4)	
Overall				
0-71	3.1 [1.7 5.5]	2.5 [1.5 4.0]	**0.6 [0.5 0.8]**[a]	×
Day 1				
0-5	1.5 [0.5 2.7]	1.3 [0.5 2.3]	**0.2 [0.0 0.5]**	×
6-11	2.2 [1.3 3.7]	1.8 [0.7 3.3]	**0.4 [0.1 0.7]**	×
12-17	3.1 [1.7 4.8]	2.2 [1.1 4.2]	**0.9 [0.5 1.3]**[a]	×
18-23	3.3 [1.8 5.9]	2.4 [1.5 3.9]	**0.9 [0.5 1.2]**[a]	×
Day 2				
24-29	3.3 [1.8 5.7]	2.8 [1.6 4.0]	**0.5 [0.1 1.1]**	×
30-35	3.7 [2.1 6.5]	2.7 [1.8 4.6]	**1.0 [0.5 1.4]**[a]	×
36-41	3.6 [2.0 6.0]	2.8 [1.7 4.3]	**0.8 [0.2 1.4]**[a]	×
42-47	3.6 [2.0 6.0]	2.9 [1.8 4.2]	**0.7 [0.2 1.1]**[a]	×
Day 3				
48-53	4.0 [2.2 6.8]	2.9 [1.9 4.4]	**1.1 [0.6 1.6]**[a]	×
54-59	4.4 [2.4 6.7]	3.2 [1.9 4.8]	**1.1 [0.4 1.6]**[a]	×
60-65	3.8 [2.3 6.0]	3.2 [2.1 4.6]	**0.6 [0.1 1.0]**	×
66-71	3.8 [2.5 5.7]	3.0 [2.4 4.7]	**0.8 [0.4 1.2]**[a]	×

Equivalence is indicated by ⇔, Non-equivalence is indicated by ×. Equivalence is a separate analysis to statistical difference. Hours where the medians are statistically different (95% CI does not cross zero) to P<0.05 are in bold.
[a]Difference remaining significant after Bonferroni correction (p<0.004).

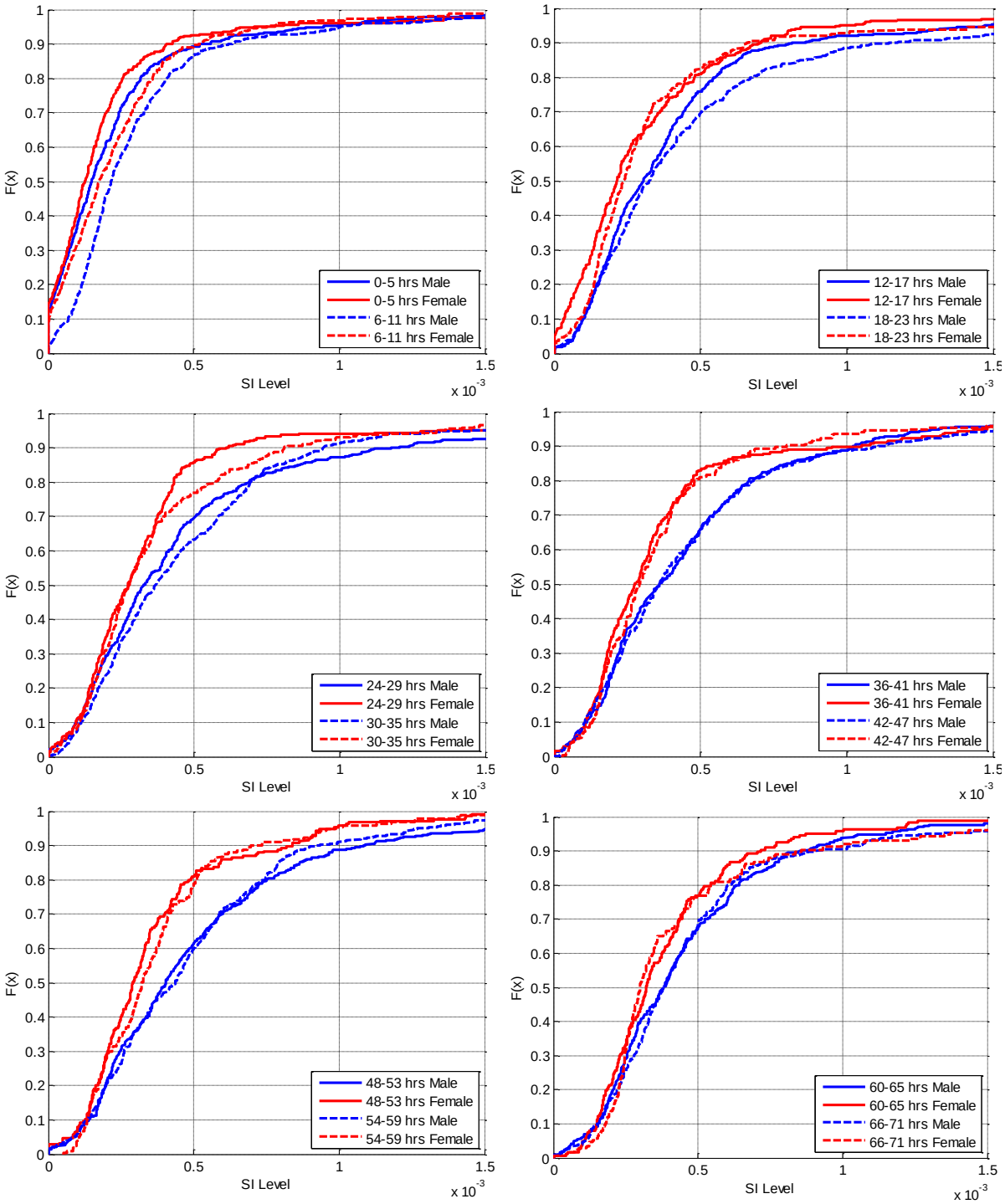

Figure 7.2 – Comparison of cumulative distribution of SI levels (L/mU/min) between male and female cohorts over 6-hour time intervals for the first 72 hours of GC.

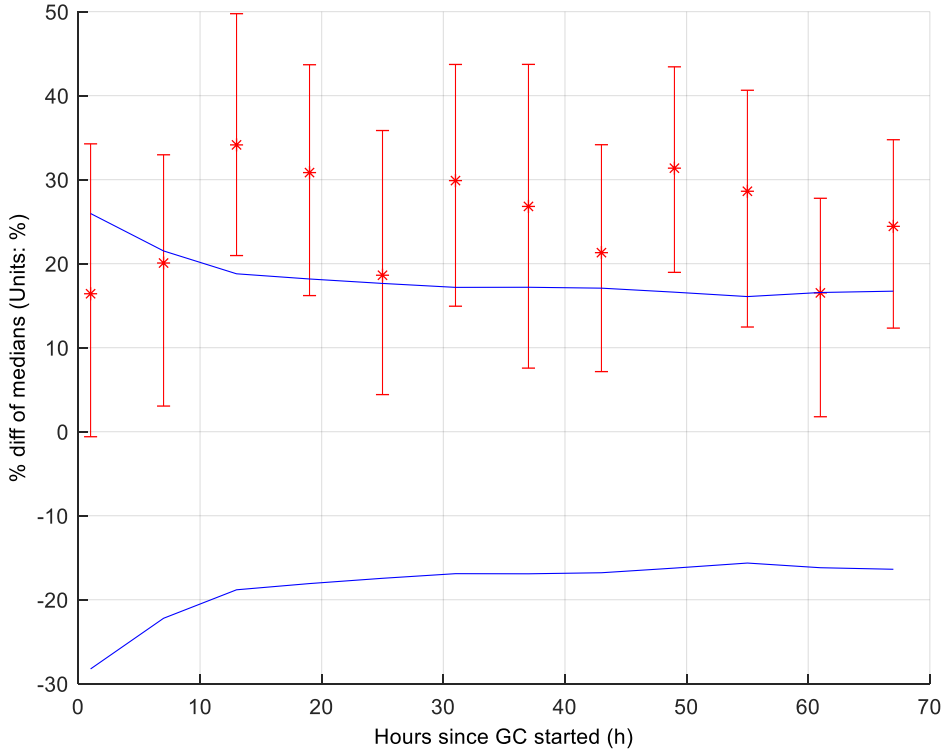

Figure 7.3 – Equivalence testing on SI for each 6-hour blocks. The blue lines give equivalence range for a typical 9.4% BG measurement error. Equivalence is accepted if the 95% CI (bars) of bootstrapped percentage difference in median SI values is within the equivalence range, and rejected otherwise.

The results of equivalence testing on SI are shown in Figure 7.3. The 95%CI percentage difference in medians between males and females is always outside the clinical equivalence range. Thus, SI levels differences between male and female are statistically different, and this difference are not clinically equivalent.

Figure 7.4 shows male and female cohorts overall %ΔSI. %ΔSI comparison for each 6-h block is presented in Table 7.3 and shown in Figure 7.5. The 95% CI of bootstrapped percentage difference in median %ΔSI levels between male and female always crosses zero, except for one 6-h block (30-35 hours). Male and female SI variability is thus not significantly different, especially if the Bonferroni correction is considered, resulting in no 6-h blocks statistically significantly different.

Furthermore, the 95% CI difference of median %ΔSI between males and females is shown in Figure 7.6 for each 6-h block in terms of equivalence. The difference is within the equivalence range for all 12 6-h blocks. Therefore, %ΔSI is not statistically significantly different for these cohorts and can also be considered equivalent.

Hypothesis and equivalence testing results for resampled (N=50) male and female sub-cohorts, with the same number of T2D patients (16%) match those of the raw, original cohort analysis. Differences in SI levels between sexes were typically significant (Figure 7.7a), and never equivalent (Figure 7.7b). Differences in %ΔSI were generally not significant (Figure 7.7c), and almost always within equivalence range (Figure 7.7d). These results confirm results from the overall population cohort analysed here.

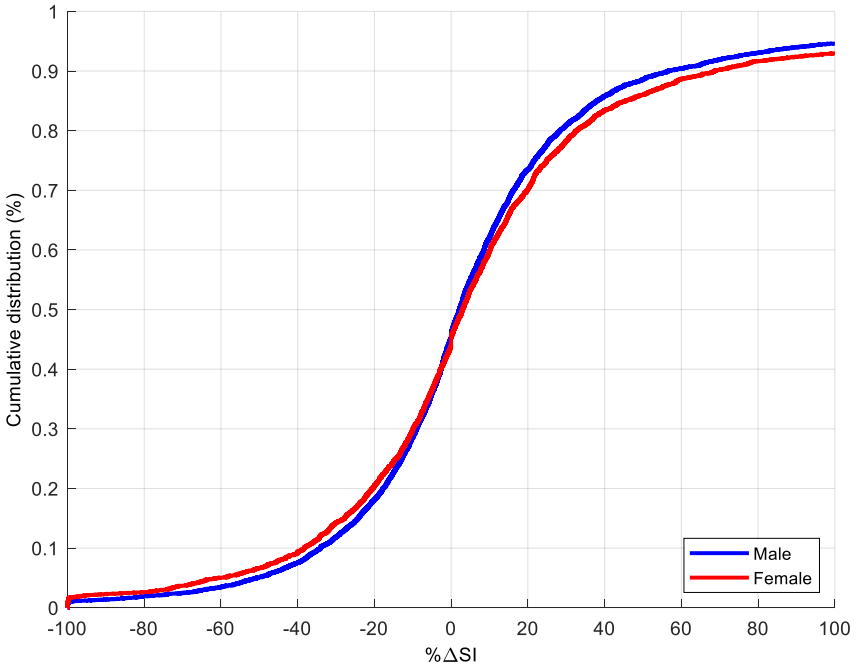

Figure 7.4 – Overall cumulative %ΔSI between male and female cohorts.

Table 7.3 – Median [IQR] %ΔSI levels comparison between male and female cohorts using 6-hour blocks

Hours	Male cohort %ΔSI	Female cohort %ΔSI	Median $\%\Delta SI_M$-$\%\Delta SI_F$ [95%CI]	
Overall				
0-71	2.2 [-17.8 21.6]	3.0 [-14.4 24.9]	-0.9 [-2.7 1.0]	⇔
Day 1				
0-5	4.5 [-23.1 61.3]	1.6 [-34.5 51.1]	8.0 [-9.7 9.8]	⇔
6-11	7.2 [-12.7 38.7]	9.9 [-15.4 42.0]	-2.8 [-9.7 4.0]	⇔
12-17	5.4 [-10.6 27.4]	4.5 [-16.3 37.6]	0.7 [-8.3 7.5]	⇔
18-23	2.9 [-15.5 24.2]	2.4 [-14.6 25.0]	0.8 [-4.7 7.1]	⇔
Day 2				
24-29	2.5 [-12.5 22.1]	4.7 [-13.0 24.9]	-2.3 [-6.9 1.4]	⇔
30-35	0.2 [-15.6 23.7]	5.6 [-12.0 24.5]	**-5.8 [-11.0 -0.7]**	⇔
36-41	1.2 [-11.4 16.2]	0.3 [-17.0 16.4]	1.1 [-3.4 6.5]	⇔
42-47	2.0 [-12.3 19.8]	0.6 [-11.9 18.2]	1.4 [-3.7 5.1]	⇔
Day 3				
48-53	2.7 [-8.6 16.3]	0.7 [-10.8 18.7]	1.6 [-2.8 5.4]	⇔
54-59	-0.8 [-15.0 13.1]	1.3 [-10.2 18.1]	-2.3 [-5.8 1.9]	⇔
60-65	1.3 [-11.0 17.5]	4.5 [-10.0 19.6]	-3.9 [-8.0 0.3]	⇔
66-71	1.9 [-9.6 13.6]	1.6 [-9.0 14.3]	-0.6 [-5.3 3.4]	⇔

Equivalence is indicated by ⇔, Non-equivalence is indicated by ×. Equivalence is a separate analysis to statistical difference. Hours where the medians are statistically different (95% CI does not cross zero) to P<0.05 are in bold. No blocks were statistically significant after the Bonferroni correction (P<0.004).

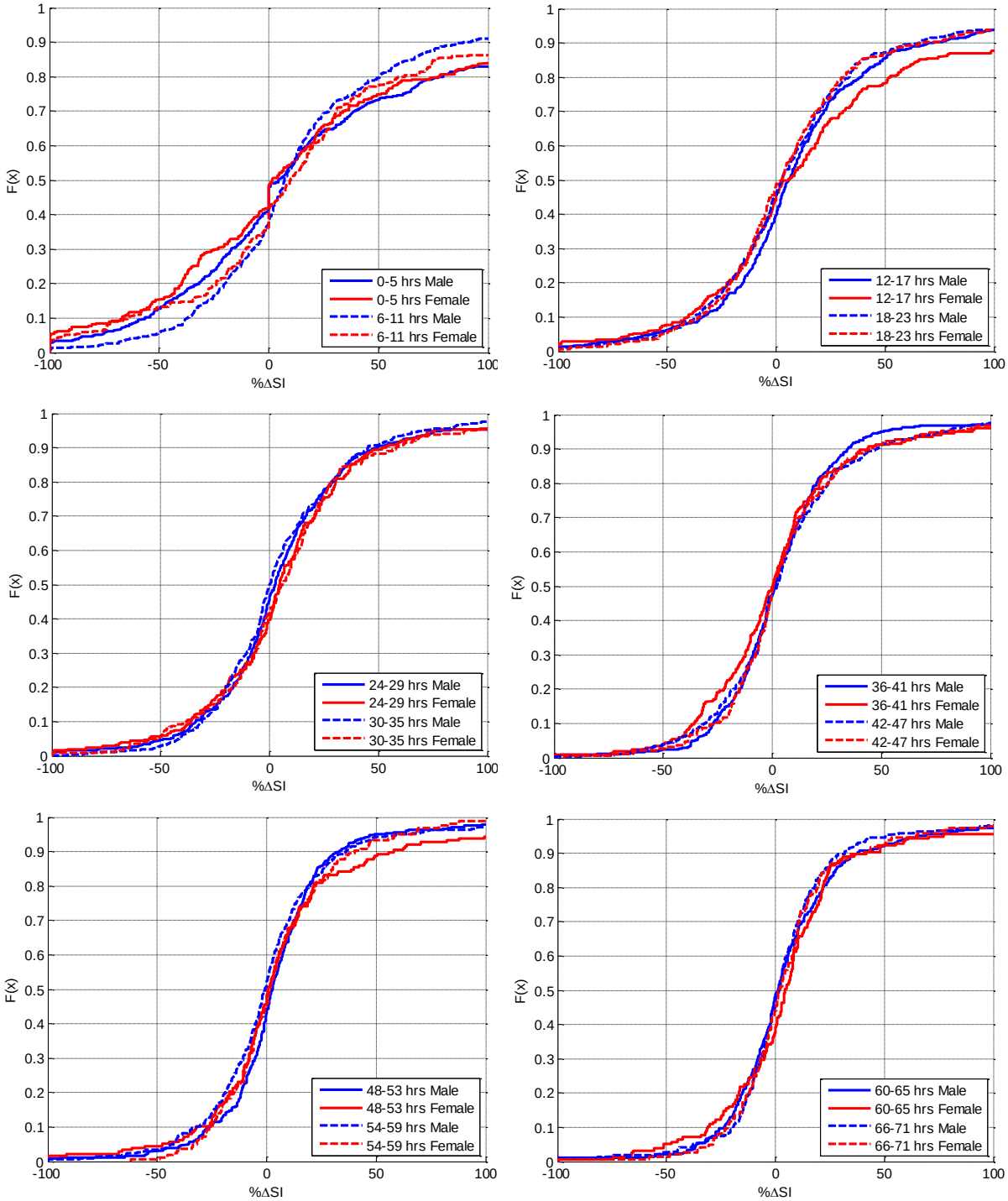

Figure 7.5 – Comparison of cumulative distribution of %ΔSI (%) levels between male and female cohorts over 6-hour time intervals for the first 72 hours of GC.

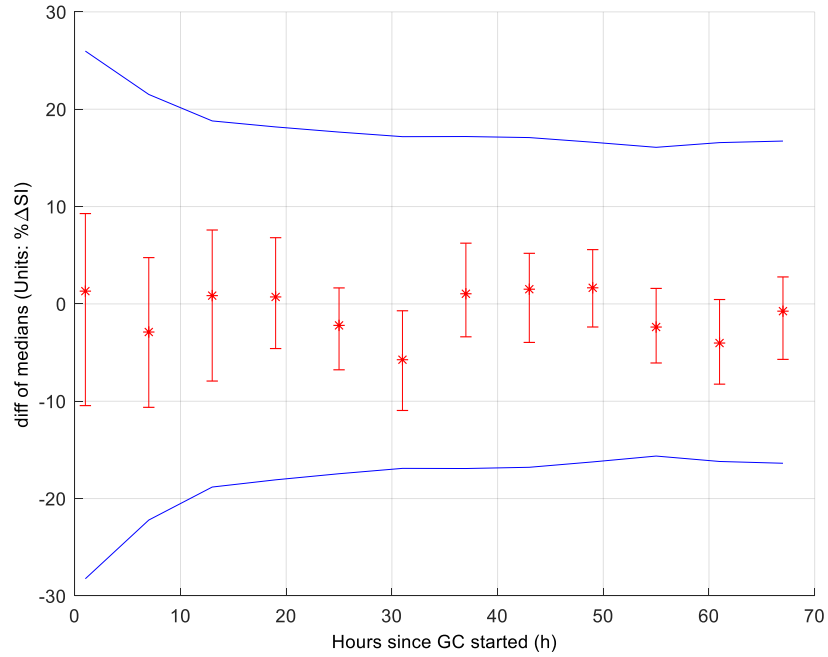

Figure 7.6 – Equivalence testing on insulin sensitivity variability (%ΔSI) for each 6-hour block. The blue lines give equivalence range for typical 9.4% BG measurement error. Equivalence is accepted if the 95% CI (bars) of bootstrapped difference in median %ΔSI values is within the equivalence range, and rejected otherwise.

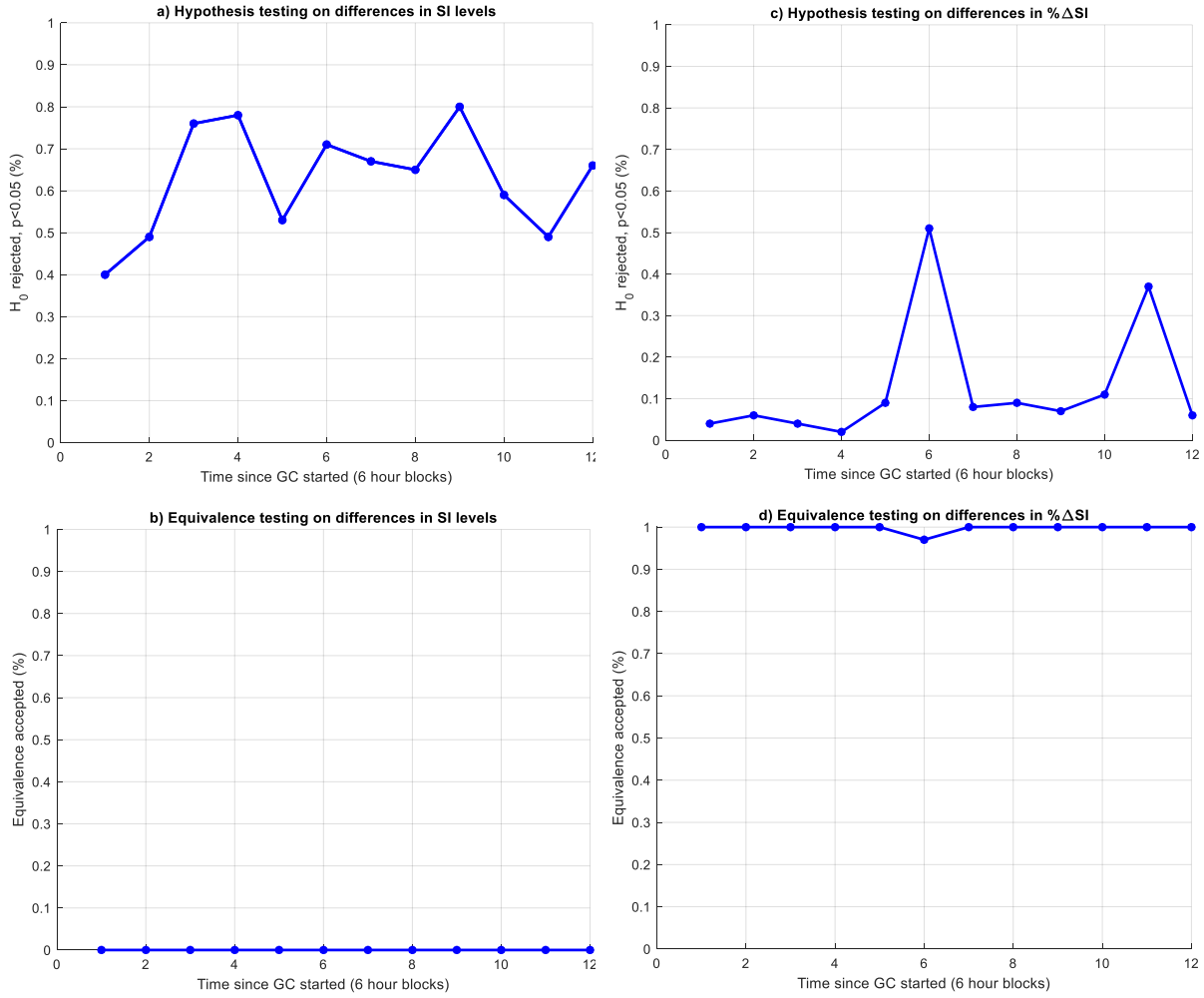

Figure 7.7 – Hypothesis and equivalence testing using 6-hour blocks from 100 resampled (N=50) male and female sub-cohorts from which 16% have T2DM.

7.4. Discussion

In this analysis, the male and female cohorts are similar in all ways (Table 7.1). Age, diabetes, severity of injury (APACHE II and SOFA scores), LOS, GC outcomes, measurement frequency, and insulin administration are all not significantly different. Only the overall cohort BG levels and the per-patient median feed administration rates achieved are statistically different (Table 7.1). However, the former is well within equivalence range considering measurement error and impact on outcomes [76, 77, 79], and can thus be considered not statistically different from a clinical perspective.

Thus, the only characteristic differentiating the two cohorts here is the consistently lower total grams of dextrose administered to the female cohort. However, this difference can arise from the typically lower caloric target for women based on lower body weight [81, 229], resulting in similar grams per kg. When nutrition is considered as the percent of the original target GF, which is consistent and based on frame size an body weight [81, 229], nutrition is not statistically different anymore (Table 7.1). Thus, overall, these two cohorts can be considered as having very similar demographic characteristics.

The results of equivalence testing on SI suggest, in addition to being statistically different, male and female median SI levels are never equivalent, clinically. In particular, it shows one would expect different clinical insulin and/or nutrition administration to account for the non-equivalence. However, the %ΔSI analysis results suggest SI variability is not statistically different and is clinically equivalent. Two observations can be taken from this set of outcomes. First, equivalent SI variability suggests both cohorts should be able to benefit from the same quality of GC, as they are equally easy/hard to control. Second, women are more insulin resistant than men (Figure 7.1, Table 7.2). In this analysis, both cohorts benefit from same GC quality (Table 7.1). All else equal, this result suggests the metabolic stress response is higher or stronger for females than for males, thus explaining this higher observed model-based insulin resistance.

No weight information was available for this cohort, but GF is calculated using the ACCP recommendation of 2000 kcal/day [229], and personalised for each patient according to age, sex, and body frame size using a standardised scale for consistency [134]. These three factors cover energy demands based on weight, sex, and age, where the first covers demand based on mass, the second accounts for differences in metabolic requirement per unit body weight for women, and the third

accounts for decreasing demand as age rises. Personalised nutrition goals can thus vary between 1025 and 2450 kcal/day over all patients.

More specifically, as noted above, nutrition was similar in %GF delivered, but higher in grams per hour for men due to their larger frame size. Thus, in Table 7.1, GF (g/h) is higher for males, as expected, reflective of their typically higher body mass. However, males and females have similar %GF, suggesting overall caloric goals per body mass are very similar across cohort, given the similar age in both groups (Table 7.1). In addition, insulin delivery was not significantly different for both males and females.

Hence, given similar %GF and total insulin administration in each group (Table 7.1), females received similar g/hr of nutrition per body weight and demand, but were given higher insulin per body mass. More explicitly, in this comparison, %GF is normalised to mass in (large) part, but insulin delivery is not. It thus confirms females require more insulin per unit of estimated body mass to remove similar amounts of glucose given per unit of estimated body mass, supporting the lower SI found for females in this analysis.

The SI metric used in this context comes from a validated physiological model and has been widely shown to correlate well with gold standard measures [136-138]. All else equal, it can be hypothesised the difference in these model-based identified SI levels would come from two main parameters in the ICING physiological model: a higher EGP for women than estimated; and/or a lower estimated insulin secretion rate. In the first case, higher EGP would suggest a stronger stress response to injury, since severity are similar across the two cohorts (Apache II and SOFA scores, Table 7.1). In the second, the lower insulin secretion would also imply a greater suppression of insulin secretion due to stress response arising from the insult compared to men. A combination is also likely, and possible, given the impact of stress response on both issues [5, 7, 13, 23, 54, 230, 231].

Until early 1990s, clinical trials were mainly conducted on men [224, 225]. Outcomes were thus biased, based on male clinical research results, leading to drug dosage for females being typically derived from average male requirements [226]. Women have clearly been under-represented in clinical trials [223], and are still severely under-represented today [222]. While their higher metabolic variability or difference

in response to treatment was seen as a potential outcome bias [228], it has been more recently stated it should be considered as a critical factor impacting outcomes [224, 226]. More specifically, some drugs, beneficial for men, may sometimes significantly increase problems in women [232] and women can have significantly different metabolic or clearance rates for drugs resulting in very different concentrations for the same dosing protocol [227]. All these points support the importance of identifying potential sex-related effects in clinical trials and care, similar to the differences shown in this study.

In particular, many clinical trials, although including both men and women, often fail to account for potential differences in drug effectiveness or safety between men and women [227]. In the context of GC, protocols are often “one size fits all” solution, lacking the ability to account for significant inter- and intra- patient variability [90, 98], where insulin dosage is similar regardless of age, body mass, or sex. However, our result shows a clear difference between males and females for insulin requirements, due to the higher insulin resistance seen in females, which would require different dosing protocols and/or a personalised approach. Model-based approaches such as in STAR, or SPRINT, and their patient-specific, risk-based approach is able to capture this variability [95], and thus, intrinsically, account for differences between patients, such as sex.

Sex differences in insulin resistance, insulin secretion, glucose effectiveness or EGP have already been shown in specific populations [219, 233-238]. The results shown in these studies sometimes contradict, but tend to say women are more sensitive to insulin than men in healthy and outpatient scenarios. In critical care patients, only one study showed a difference, demonstrating, in opposition to the above studies, higher insulin secretion and thus greater resistance in preterm girls compared to preterm boys in the neonatal ICU [219, 220]. These NICU results would not necessarily be expected to extend to adults, but the results presented show the same bias in adult ICU cohorts, suggesting a different in metabolic stress response at these two extremes of age and development.

It is important to note, many studies have analysed differences in mortality outcomes, treatment effort, or other factors between sexes in ICU patients. However, these studies often contradict. Some showed higher mortality in women [239-241], but others did not [242, 243]. The differences between sexes are thus still not completely understood in ICU, although present [244], showing the importance of assessing the related potential implications, as done in this study.

To the authors' knowledge, there were no studies analysing EGP or insulin secretion between sexes in adult ICU populations, which could differ in many ways due to their acute metabolic conditions. This study thus appears to be the first study suggesting women could be more resistant to insulin compared to men in this cohort, and that this outcome could be due to their potential greater response to insult induced stress.

Despite the relatively small cohort size considered, an advantage of this study is the quality of the data and its detailed nutrition and insulin input information. In addition, the cohort is smaller because ensuring consistent start of GC from ICU admission of <12 hours eliminate bias due to patients being considered at different point in the evolution of stress response. Hence, the smaller cohort, while still providing sizeable data, is a result of eliminating a potential bias in this time-based analysis.

The observations made rely on the identification of the SI parameter using a mathematical model, where inaccuracies could lead to bias. However, the ICING model typically performs well in the clinical ranges observed here, suggesting low inaccuracy. Furthermore, it has been validated in extensive clinical use [81, 87, 93, 156, 245]. The validity of SI metric and its use in the context of this thesis is discussed in Chapter 6.

This study does rely on retrospective data from a single centre study, which could limit the clinical impact of these results, though, in contrast, the data reflect a generalised cohort of patients across multiple years of clinical practice. In addition, the lack of reported demographic information, such as weight, and body mass index, are a limitation to consider the caloric goals per body mass similar across cohort in this analysis, which were only inferred in this study due to their use in setting GF rates. Finally, only sex and known diabetes mellitus have been considered in this analysis, while other confounders, such as ethnicity, could potentially influence the results. There might also be patients with unknown diabetes in the cohort, where measures of HbA1c could have helped to clearly identify these patients, but are not available here.

7.5. Summary

This study compared identified SI and %ΔSI across male and female cohorts using hypothesis and equivalence testing. SI was shown statistically significantly lower for females and this difference is clinically not equivalent to males. However, %ΔSI between males and females was not statistically different, and clinically equivalent. These results strongly suggest females may have stronger metabolic stress response than men. These results also suggest higher insulin requirements for females, while equal safety and efficacy should be able to be achieved for both cohorts, as reflected in the equivalent variability. Future GC RCTs should thus also consider randomising and analysing male-female subgroups for differences in primary and secondary outcomes.

These results thus suggest inter-patient variability is clinically different, but intra-patient variability is clinically equivalent. In turns, these results are similar to Chapter 6 comparing survivors and non-survivors. Hence, it can be deduced inter-patient variability is likely always different while intra-patient variability is always equivalent. In a published retrospective analysis comparing different cohorts of patients from New Zealand and Belgium, intra-patient variability was found significantly different but not intra-patient variability [86], supporting once again these observations. However, further analysis comparing other different sub-groups based on other specific demographics should be undertaken to generalise this conclusion.

Overall, Chapters 5-7 emphasise the importance to provide safe, effective control for (nearly) all patients, and highlight the key role of both inter- and intra- patient variability. Since the quality of control depends upon the ability to account for variability, the better this variability can be characterised, the likely better GC outcomes will be.

Chapter 8: Characterising Variability – A First Approach to Improve Predictions of Intra-Patient Variability

Over the previous chapters, inter- and intra- patient variability was shown to directly impact GC safety and efficacy. More specifically, inter-patient variability determines patient-specific need for higher or lower insulin doses to reach the target, while intra-patient variability reflects the risks associated with a given dose of insulin. Thus, while the former provides information of current patient-specific metabolic state to adjust treatment, the latter is what makes GC hard to achieve safely, as a relatively sudden change during a treatment interval could lead to hypoglycaemia.

In Chapter 6 and Chapter 7, analysis of SI levels and variability showed inter-patient was not clinically equivalent when comparing sexes and clinical outcome, but intra-patient variability was equivalent. Hence, these results support the use of a population-based stochastic model in STAR. They also suggest this stochastic model, predicting intra-patient variability, to be critical for safety and GC success.

The goal of this chapter is thus to determine if prediction of intra-patient variability can be improved and better characterised using more information on prior SI evolution. Doing so would enable more personalised control in the context of STAR, as these predictions are directly used to determine treatment and assess risks, using its unique risk-based dosing approach (Chapter 3). In essence, can intra-patient variability be made more patient-specific and less (full) cohort specific?

This chapter presents results published in [246] and [126].

8.1. Introduction

STAR identifies patient-specific SI, characterising inter-patient variability, and predicts its future variability, characterising intra-patient variability [95, 96, 123, 149, 151]. While the identification of patient-specific SI levels is important to capture current patient metabolic state, a good prediction of its future variability is critical. Better forward prediction of SI allows better characterisation of future metabolic variability, thus improving patient-specific GC without compromising safety. In the context of STAR, improved, more patient-specific, prediction of future SI levels could have a significant impact on treatment selection, better mitigating hypoglycaemic risk (Chapter 3).

The stochastic model currently used by STAR forecasts future SI (SI_{n+1}) distributions based on the identified current SI value (SI_n) [124, 125]. This 2D stochastic model was originally constructed using clinical data and kernel density methods [124, 125]. A Markov process is used, where outcome SI_{n+1} only depends on input SI_n [125], and the future SI_{n+1} distribution can be used to derive likely future BG distributions for a specific insulin and nutrition intervention [95, 96, 123, 149, 151].

More specifically, this chapter aims to determine if the prediction power of the existing 2D stochastic approach can be improved by also using the recent changes in SI, as well as its current level, as input parameters for forward prediction of outcome SI_{n+1}. The new 3D stochastic model will now predict future SI_{n+1} based on current SI_n and the percentage change in SI from SI_{n-1} to SI_n. The old 2D the new 3D stochastic models are compared to assess the new model's ability to tighten SI prediction ranges for tighter forward prediction of future BG. Narrower future SI prediction ranges enable more targeted insulin dosing for those patients who are more stable, and vice-versa for less stable patients. By better characterising variability, this analysis also assesses whether more stable patients have lower future metabolic variability.

8.2. Methods

8.2.1. Patients and Cohorts

This study uses data from 3 clinical ICU data cohorts totalling 819 GC episodes (606 patients) and 68629 hours of treatment [81, 87]. The SPRINT, STAR Christchurch, and STAR Gyula cohorts have been presented in Chapter 4.

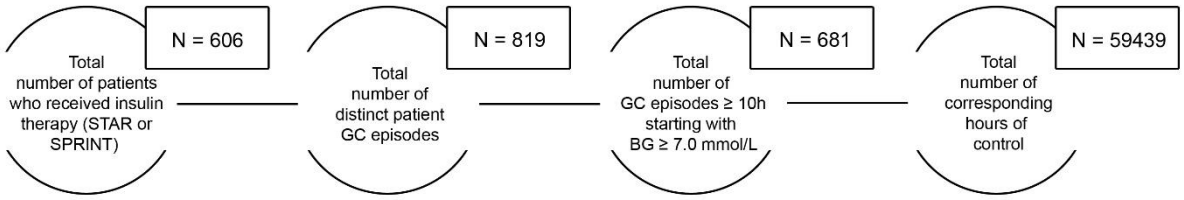

Figure 8.1 – GC episodes selection from the original 606 patients (819 different GC episodes).

From the original 819 episodes, only 681 episodes ≥ 10 hours and with initial BG ≥ 7 mmol/L are considered (Figure 8.1), corresponding to 59439 hours of control. These criteria ensure the exclusion of patient data with very short GC episodes, and thus low BG measurement numbers, or uncommonly low starting BG values, which are likely less reflective of general metabolism dynamics [87]. SI is identified hourly for each patient using integral-based fitting methods [130, 131] and a total of 58539, 57840, and 57141 data triplets ($\%\Delta SI_n$, SI_n, SI_{n+i}) for i = 1, 2, and 3 hours forward predictions, respectively, are created.

8.2.2. Analysis

The existing 2D stochastic model uses the input SI_n to determine the outcome distribution of SI_{n+1} [124, 125]. This study builds a new 3D model to determine the outcome distribution of SI_{n+1} based on input of patient-specific current metabolic state, SI_n, and SI variability to current time, $\%\Delta SI_n$ (Equation (6.1). Thus, more stable patients would have low or near zero $\%\Delta SI_n$ values. This choice thus further delineates sub-cohorts of patients by their metabolic variability at a given SI_n level.

The data triplets were binned with bin increment sizes of $\%\Delta SI$ = ±10% and SI_n = 0.5e-4, based on BG measurement error on SI (Appendix II) [135]. These bins are limited to a range of $\%\Delta SI$ = [-100%, 200%] and the 1^{st}-99^{th} percentile range in identified SI ([1.0e-7, 2.1e-3] L/mU/min) values, bringing the total number of triplets considered to 97% of the original data triplets, as those few outside these ranges are excluded.

The minimum number of data points required for adequate data density in each bin was arbitrarily defined to be 100 data triplets to ensure any distributions were not influenced by outliers. To improve data density and smooth model extremes, bins not meeting this criterion are summed together along

the %ΔSI axis at the same SI_n level, allowing data triplets to influence neighbouring bins where there is insufficient data density. The summation process is described below, and an example is shown in Figure 8.2.

Starting from the bin centred at 0% and going down, and at the bin centred at 10% and going up:

1. Check data density.
    a. If the number of data triplets is >=100, move to the next bin.
    b. If the number of triplets < 100 add the triplets from the 'outer' adjacent bin(s) until data density is reached. If summation of bins still results in a failure to reach the data density, stop here.
2. Repeat step 1 until the model limits are reached.

Bin along %ΔSI	# triplets in bin for a given SI_n	# triplets in bin after summing bins together.
[-45, 55] %	0	-
[-35, 45] %	2	-
[-25, 35] %	10	-
[15, 25] %	25	37
[5, 15] %	80	105
[-5, 5] %	150	150
[-15, -5] %	102	102
[-25, -15] %	10	11
[-35, -25] %	1	-

Figure 8.2 – Example of data density before and after merging bin process, where the joined lines show the bins merged to create the fined surface.

The 5th, 50th, and 95th percentiles of SI_{n+1} are computed for each bin. These values define a non-parametric 5th – 95th percentile distribution range and median likely outcome of future variation of SI values based on a specific current SIn level and the previous change $\%\Delta SI_n$. They are interpolated linearly between bins to give a percentile surface.

To compare this new 3D model to the previous 2D model, the percentage change in the 5th, 50th, and 95th percentiles are analysed, as well as the percentage change in the 5th – 95th percentile prediction

range width used in STAR GC to select insulin and nutrition doses. This width effectively determines the outcome BG range for a given intervention. Hence, a narrower width allows better prediction performance.

Hence, this analysis aims to identify regions of the model that are either conservative or have higher risk of hyper- and hypo- glycaemia. A narrowing of the 5^{th} – 95^{th} percentile range in the 3D model compared to the 2D model would allow the STAR controller to dose insulin more aggressively as outcome likelihoods are tighter. In contrast, wider 5^{th} – 95^{th} percentile range than in the 2D stochastic model indicates increased risks, when using the 2D model, and insulin is more conservatively dosed based on this new information.

8.2.3. Validation

A preliminary validation of the model is carried out by assessing its ability to predict SI in clinical data episodes longer than 24 hours. The two models are compared by evaluating the per-patient percentage of SI outcomes falling into the model-predicted 5^{th} – 95^{th} the 25^{th} – 75^{th} percentile ranges for each model. Ideally, all episodes have exactly 90% and 50% within these ranges indicating a cohort derived model that is also perfect for each patient episode. Overall, this metric gives a measure of the model's ability to capture clinically observed patient-specific changes in SI [247], as well as quantifying the cohort-derived model's level of patient-specificity. The overall goal of this comparison is a more patient-specific stochastic model.

To validate model consistency in a cohort different from which it was developed, cross validation simulations are used. A 3D model was built from SI traces from a randomly selected group of episodes comprising of 70% of all episodes, and tested on the remaining 30%. Per-patient percent time in the 25^{th} – 50^{th} and 5^{th} – 95^{th} percentile ranges are computed, as well as the ratio between the widths of the 5^{th} – 95^{th} percentile prediction ranges for both models. This process is repeated 50 times using episodes with at least 24 hours of clinical GC data. Significant variability in results indicates a model based on too little data and/or dominated by selected patients or episodes. Consistent results indicate the model is built of enough data and/or is not skewed by outlying data. This analysis thus assesses model robustness to its underlying data.

8.3. Results

8.3.1. Forward Prediction Performance of the 3D Model

The number of data triplets ($\%\Delta SI_n, SI_n, SI_{n+1}$) per bin is shown in Figure 8.3**,** before and after bin merging to improve data density. The triangular shape of binned data suggests greater variability at lower SI, where changes constitute a larger percentage change relative to absolute SI value. Yellow areas represent bins with enough data density (at least 100 data triplets) and thus the bins used to build the model.

The new interpolated 3D model is shown in Figure 8.4 and compared to the original stochastic model (green) for the 5th (a) and 95th (b) percentiles. The previous 2D model forms a plane in the new 3D model space as it is constant across all %ΔSI. Where the new 3D model sits above the 95th percentile or below the 5th percentile planes indicates where the 2D model was too narrow and thus not conservative or safe enough. The reverse case of above the 5th and/or below the 95th percentile indicates the 2D model was over conservative and the 3D model 5th – 95th percentile prediction range is narrower.

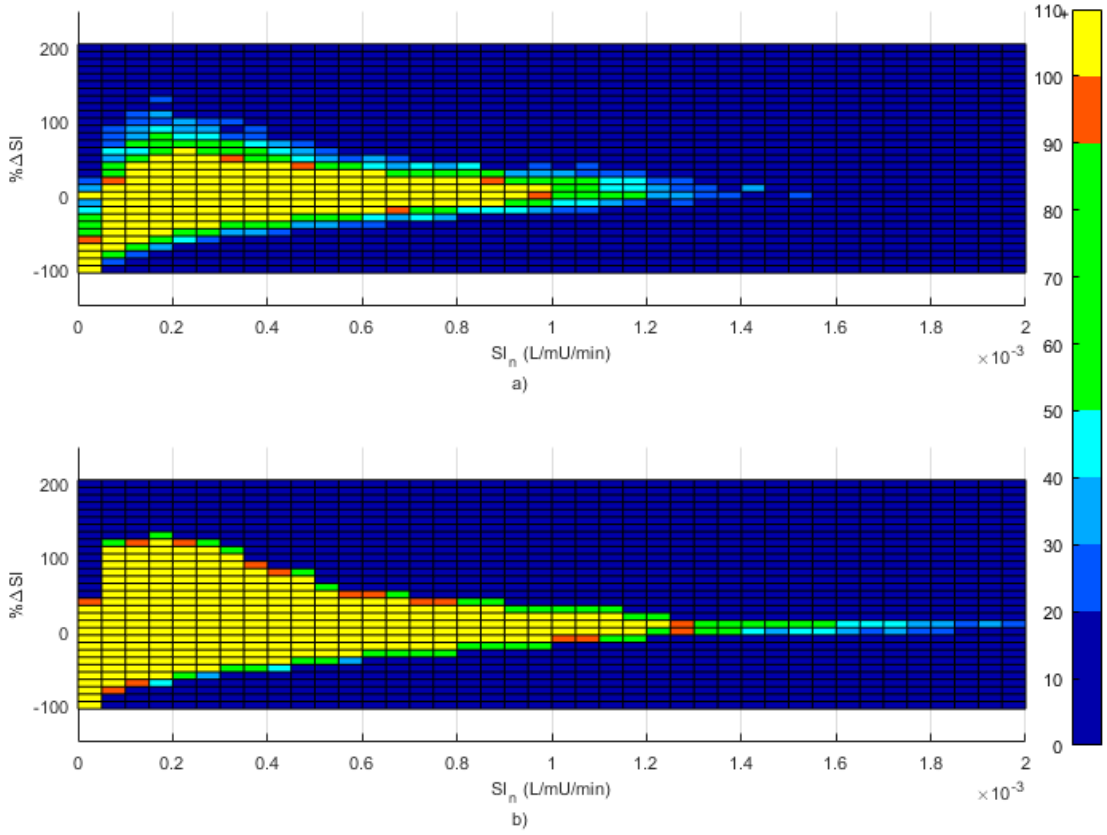

Figure 8.3 – Number of data triplets per bin before (a) and after (b) merging side bins along the y-axis. Bins in yellow reach minimum data density and will be used to build the 3D model.

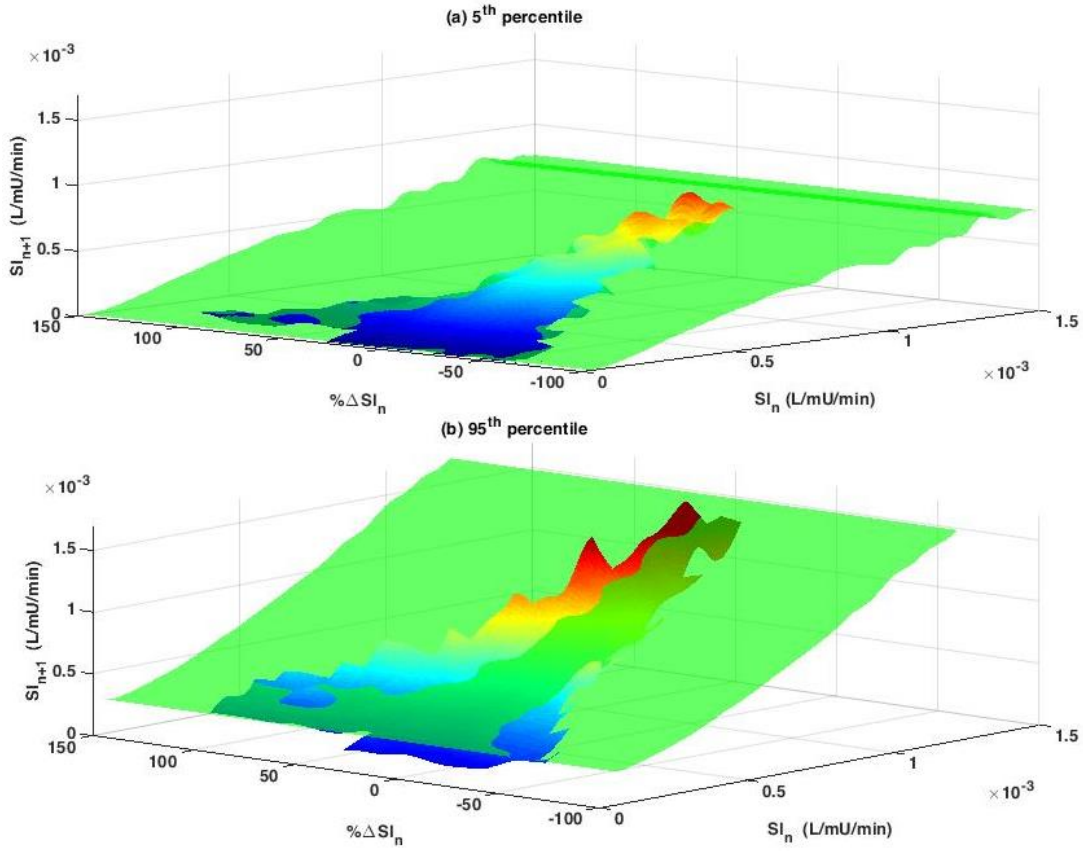

Figure 8.4 – Comparison between the new 3D model (colour) and the original 2D model (green) for the 5th (a) and 95th (b) percentiles.

The percentage change in the 5th, 50th and 95th percentiles prediction of SI_{n+1} is shown in Figure 8.5. Two main regions can be identified:

- A conservative region, mainly between $\%\Delta SI_n = \pm 25\%$. The 5th percentile is higher than the previous 2D model, while the 95th is lower, regardless of SI_n, describing thus a narrower 5th – 95th percentile range in the forward prediction of SI_{n+1}. This region represents 77.2% of the data triplets.

- A non-conservative region outside $\%\Delta SI_n = \pm 25\%$. The 5th percentile is lower and the 95th is higher, indicating higher resulting risks of hyper- and hypo- glycaemia than predicted by the 2D model.

- Summarising Figure 8.5, the percentage change in the 5th – 95th percentile prediction range is shown in Figure 8.6. In the 2D model's conservative region, a significant decrease of ~25-40% in this range is observed for the 3D model, suggesting the new model allows more aggressive insulin treatment than the previous model and thus provides improved information for dose selection. For the non-conservative region, increases in SI of up to 80% or more are observed,

allowing the model to more safely predict and cope with large changes in SI in regions of high metabolic variability (high %ΔSI). The reduced width green region contains 79.7% of the total data triplets, indicating a significant number of hours with over-conservative treatment selection.

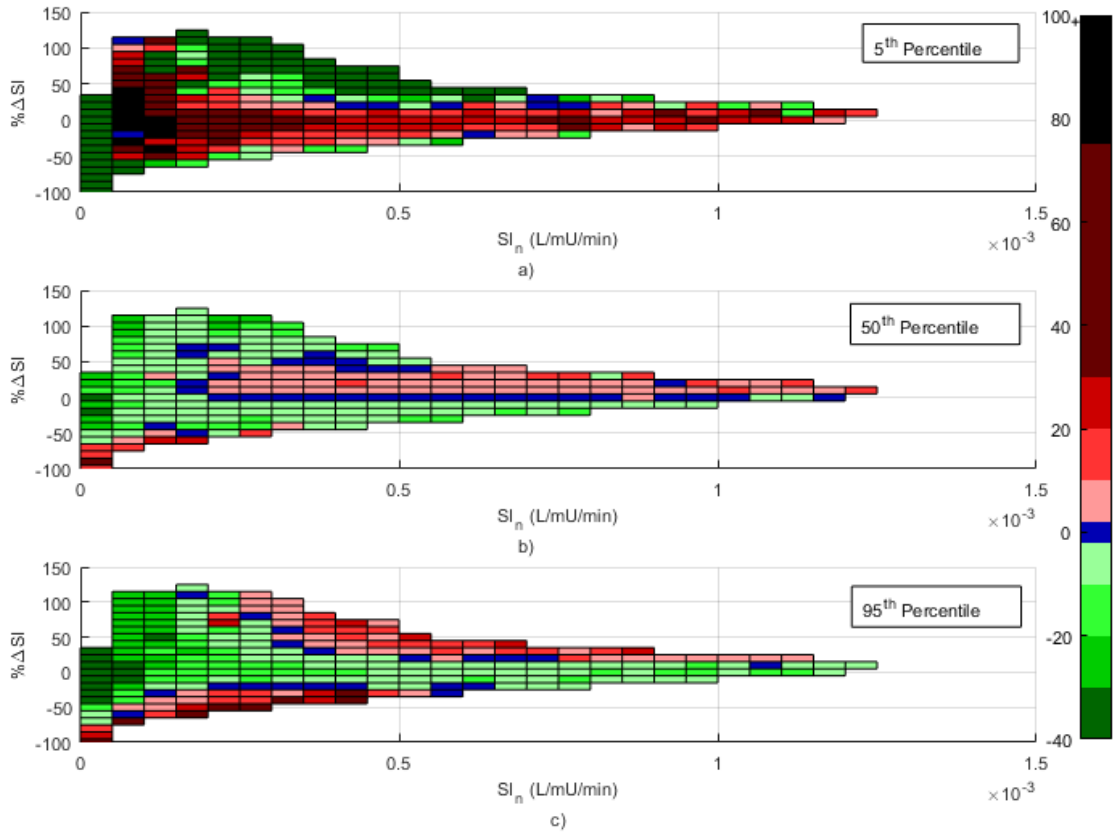

Figure 8.5 – Percentage change in the 5th (a), 50th (b), and 95th (c) percentiles between the original 2D and the new 3D stochastic models.

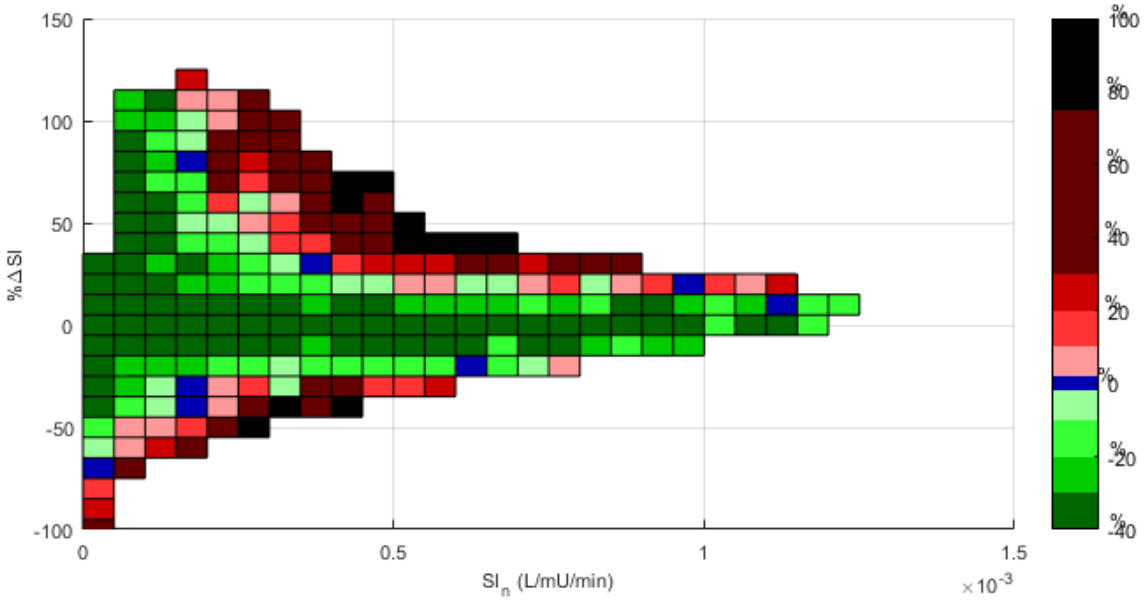

Figure 8.6 – Percentage change in the width of the 5th-95th percentile prediction range between the 3D and 2D models. Green and red areas suggest over and under conservative behaviour respectively within the 2D model.

8.3.2. Self-Validation

The predictive power of the new model is tested on 587 episodes of minimum 24 hours. In total, 90.4% of SI_{n+1} predictions were within the 5^{th} – 95^{th} percentile prediction range, and 52.6% fell within the 25^{th} – 75^{th} percentile range, which is very close to the expected values of 90% and 50%, respectively. The 5^{th} – 95^{th} percentile interval is more critical due to its use in dosing. The larger error in the 25^{th} - 75^{th} percentile interval versus the 5^{th} – 95^{th} percentile interval indicates a small mismatch in the distribution shapes across SI_n.

Figure 8.7 presents a histogram of per-patient percentage forward prediction in the 25^{th}-75^{th} and 5^{th}-95^{th} percentile bands, showing the accuracy at a per-patient level, rather than an overall cohort level. Table 8.1 shows the corresponding per-patient median [IQR] percentage time in these bands. The per-patient median [IQR] percentage time in the 25^{th} – 75^{th} percentile prediction range is higher in the previous 2D model (60.3% [47.8%, 71.5%] vs. 51.2% [42.9%, 59.2%]), while the percentage time in the 5th – 95^{th} percentile prediction range is similar (93.6% [85.7%, 97.3%] vs. 90.7% [84.4%, 94.6%]) between the 2D and 3D models. However, as seen in Figure 8.7, the per-patient distributions are tighter to the ideal values (50% and 90%) for the 3D model, reducing over-conservatism (and risk) per-patient.

More importantly, there was a significant reduction in the width of the 5^{th}-95^{th} percentile ranges for each patient, with the 3D model reducing this width by median 28.9% [21.6%, 33.0%] per bin. These results suggest the new 3D model is able to account for changes in SI equally well in comparison to the 2D model, but with significantly narrowed prediction range for many hours of care. This outcome should allow safe application of more aggressive insulin treatments for more stable patients. An example comparison of the predictive power of the two models is shown in Figure 8.8.

Table 8.1 – Per-patient predictive power comparison between old and new stochastic models.

	2D Model	**3D model**
Median per-patient % prediction within 25^{th}-75^{th} percentile range	60.3% [47.8%, 71.5%]	51.2% [42.9%, 59.2%]
Median per-patient % prediction within 5^{th}-95^{th} percentile range	93.6% [85.7%, 97.3%]	90.7% [84.4%, 94.6%]
Median per-patient % reduction in 5^{th}-95^{th} percentile range width	28.9% [21.6%, 33.0%]	

Results are given as median [IQR].

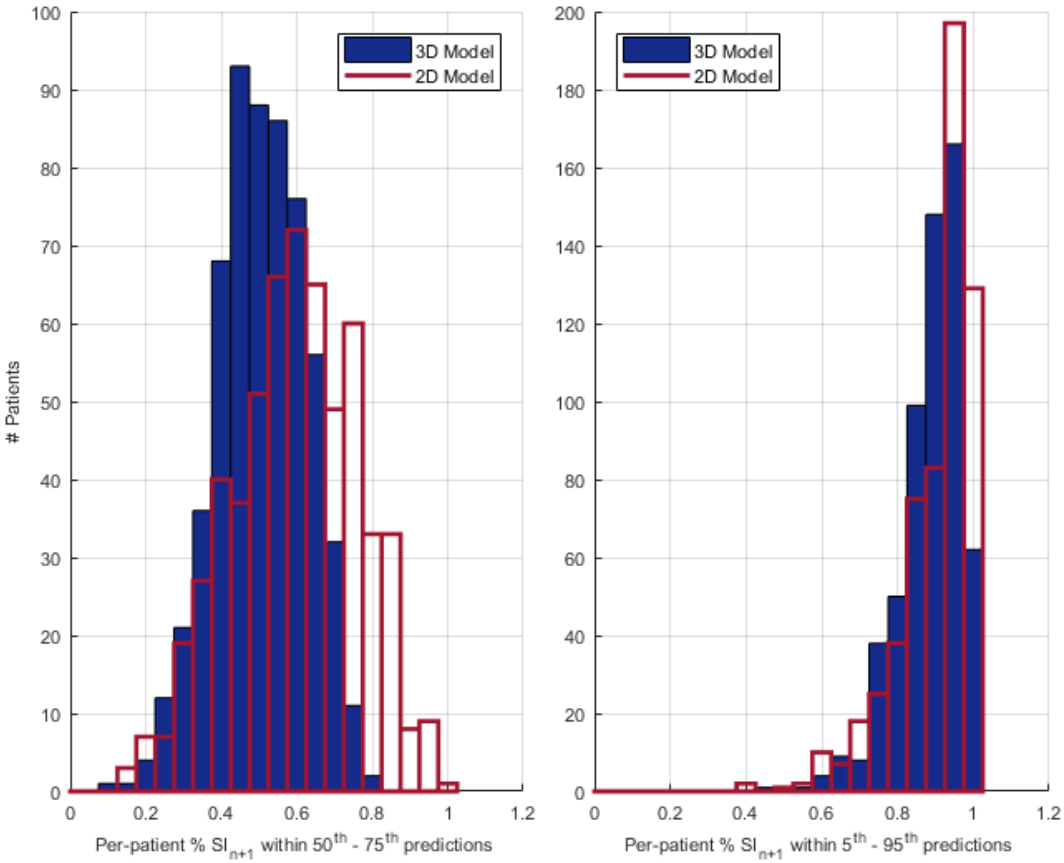

Figure 8.7 – Per-patient predictive power within the 25th-75th percentile prediction range (left) and within the 5th-95th percentile prediction range (right) of the new 3D (blue) and old 2D (red) models.

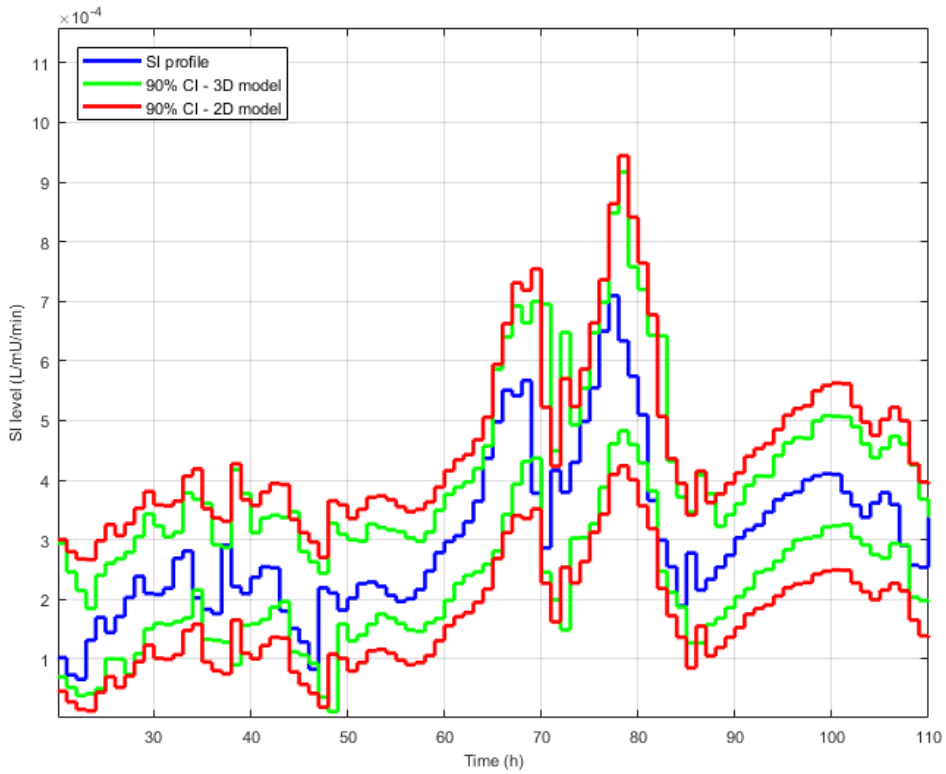

Figure 8.8 – Excerpt from a patient showing fitted SI (blue) as well as 5th-95th percentile prediction range for the new 3D (green) and the old 2D (red) models. The new model predictive range is generally narrower than the old model.

Table 8.2 compares prediction outcomes for patients who had increased and decreased time in prediction ranges, respectively. A total of 101 episodes increased percentage time in the 5th – 95th percentile prediction range by ~5% (87.3% [80.0%, 92.7%] vs. 82.3% [71.8%, 88.9%] for the 2D and 3D models, respectively). Conversely, for the remaining 486 episodes, the new model shows slightly lower performance (91.1% [85.7%, 95.0%] vs. 94.6% [88.9%, 97.6%]). However, the percentage time in range for these patients has been brought closer to the intended 90%, ensuring the 3D model treats patients more consistently across the cohort.

Table 8.2 – Per-patient predictive power comparison between old and new stochastic models for two groups: 101 patients whose % prediction increased with the new model and 486 patients whose % prediction decreased.

	2D Model	3D model
Median % prediction within 5th-95th percentile range for 101 patients who increased time in range.	82.3% [71.8%, 88.9%]	87.3% [80.0%, 92.7%]
Median % prediction within 5th-95th percentile range for 486 patients who decreased time in range.	94.6% [88.9%, 97.6%]	91.13% [85.7%, 95.0%]

Results are given as median [IQR].

8.3.3. Cross validation

Cross-validation with 70% of the data used to build the model was carried out using patient episodes 24 hours or longer, and results are shown in Table 8.3. Compared to the original 2D stochastic model, the new 3D model has consistent, 12% absolute lower median [IQR] percentage forward prediction in the 25th-75th and 5th-95th percentiles ranges (3D: 51.8% [51.5%, 52.1%] vs. 2D: 63.1% [62.8%, 63.4%] and 89.8% [89.6%, 90.0%] vs. 92.5% [92.4%, 92.6%]). Additionally, the 5th-95th percentile range width from 2D to 3D model is reduced by median 30.8% [30.5%, 31.1%]. These results suggest both models generalise well to other ICU patients when developed from an independent, but similar cohort of patients, matching similar tests across cohorts [87].

Table 8.3 – Cross-validation per-patient results for old 2D and new 3D stochastic models, on all SI values from episodes of minimum 24 hours. Results are given as median [IQR].

	2D Model	3D model
Median % prediction within 25th-75th percentile range	63.1% [62.8%, 63.4%]	51.8% [51.5%, 52.1%]
Median % prediction within 5th-95th percentile range	92.6% [92.5%, 92.7%]	89.7% [89.6%, 90.0%]
Median % reduction of the 5th-95th percentile range width from 2D to 3D model	30.8% [30.5%, 31.1%]	

Results are given as median [IQR].

8.4. Discussion

8.4.1. Main results

Forecasting changes in SI underpins the ability of STAR to respond in a patient-specific manner to potential future changes in patient GC requirements, resulting in safe and effective, risk-based GC. If the distributions of forecast likely SI changes are narrower, then control can be further improved, with tighter control in more stable patients, and better avoidance of hypoglycaemia in patients exhibiting high glycaemic variability. In this study, current probabilistic forecasting methods have been extended to include the change in SI (%ΔSI) as an input predictor for future SI alongside current SI.

The previous 2D stochastic model is shown to be conservative for ~77% of the data, where the $\%\Delta SI_n$ is within ± 25% change. While conservatism results in wider prediction ranges in likely BG outcomes, thus further reducing the risk of hypoglycaemia, it also inhibits the controllers' ability to reduce BG to the normal range using more aggressive control dosing. This issue particularly affects patients who tend to remain stable, but are mildly hyperglycaemic as a result. Hence, such conservatism, while safe, has a potentially negative clinical impact, as well.

Compared to the 2D stochastic model, an over-conservative region equally implies an under-conservative region. This under-conservative region means there are less stable patients, outside %ΔSI = ±25%, who have increased hypoglycaemic risk from relatively over-aggressive dosing, resulting from prediction bands that are too narrow. This trade off of conservatism is seen in Figure 8.5 and Figure 8.6, and offers unintended increased risk for these patients in using the 2D model.

The new model utilises change, %ΔSI, as an additional model input to better predict a more patient-specific future SI_{n+1}, with narrowed prediction ranges for 77% of hours, and overall similar ability to meet the expected 90% of SI outcomes within the 5^{th}-95^{th} percentile prediction range. The new model is thus more patient-specific, and better predicts likely BG outcomes. These results should translate into more aggressive insulin dosing where patients are more stable and SI outcomes are more certain, and less aggressive, lower insulin doses in patients who are more variable. Greater patient-specificity also reduces risk for more variable patients. This model could thus lead to more personalised, tighter, and

less variable control for all patients, with greater safety from hypoglycaemia, and thus improved outcomes [78, 79].

Both the self-validation and the cross-validation tests have shown the predictive power of this new model to be more consistent, with closer to 90% of forward prediction of SI within the 5^{th}-95^{th} percentile range. These results align well with the expected 90% of SI outcomes falling within this prediction range. However, in comparison to the previous 2D model, the new 3D model achieves this performance with overall narrower and tighter prediction distributions for 77% of hours. A median reduction in the 5^{th}-95^{th} percentile prediction range width of approximately 30% was achieved with the 3D model, indicating the 3D model is better able to predict future SI outcomes, and thus safely allowing significantly increased insulin dosing.

The new 3D model thus treats patients more consistently across the patient cohort. Previously, with the 2D model, some patients within the cohort had much more than 90% of their SI outcomes within the 5^{th} – 95^{th} percentile prediction range, and others much less (Figure 8.7). While greater conservatism is advantageous for avoiding extreme BG outcomes, it also implies an over conservatism and inability of the model to meet its design specifications (90% of SI within the 5^{th} -95^{th} percentile range). Equally, it prevents aggressive dosing and better control where it could be warranted for specific patients, and is thus less patient-specific than the new 3D model. Given the improved 5^{th} - 95^{th} percentile range performance of the new 3D model, it is clear it is better able to consider patients more consistently across the cohort with the added %ΔSI model input.

Cross simulation tested the ability of the model to predict SI in patients not used to build the model. Cross validation results were very consistent and close to expected values from the whole-cohort analysis, suggesting the model would generalise well to ICU patients from different protocols and/or units. It also indicates the model is not dominated by smaller outlying subsets of patients.

This result also reflects previous results where similar and consistent SI variability was seen across 3 different ICU cohorts in 3 different countries [87, 128]. This previous analysis uses slightly different patient cohorts from those presented here, and includes an additional cohort not used here. Thus, these

results suggest the model generalises well to other ICU patients and cohorts, when the model is built from an independent, but overall similar cohorts of mixed medical ICU patients.

8.4.2. Limitations

One limitation of this analysis is the binning process, which creates a discretized 3D space, where percentile surfaces were linearly interpolated between bins. Kernel-density methods, used to build the previous 2D stochastic model [125], could also be applied to generate smoother prediction percentile surfaces with respect to data density. However, this approach introduces assumptions around the shape of data density distributions and the effect of surrounding data on percentile surfaces. In this analysis, bin sizes were chosen to balance increased resolution against sufficient data density, as well as to first prove the concept.

The bin width of 10% in %ΔSI was chosen based on a previous analysis, which assessed the impact of BG measurement error on SI [135]. Given over 68,000 hours of patient data are used to build the new 3D model, the percentile surfaces of the full model are likely to be sufficiently reflective of SI dynamics, as data density was typically sufficient across typical areas of interest. Thus, limitations due to data size and density are likely to be minimal.

A further possible limitation of the model is that ~7% of data points fell outside the model range, having been discarded as outliers in regions of insufficient data density. These unusually large changes in SI or extremely high SI levels are thus not included in the model, and may reflect inaccuracies in data recording or patient-specific deviations from model-dynamics. It is also possible these points highlight times of extreme glycaemic change or measurement error, where the best clinical practice could be to discontinue insulin for an hour and come back and re-measure. In essence, for these outlying, potentially unexplained events, discontinuing insulin for a short period is a safe course of action. Within the STAR framework, the clinical usage would be to utilise the original proven 2D model, using the safest and most conservative intervention [87, 95, 128]. Hence, points falling in this range could be used to warn of outlying events that might not otherwise be noted, and be used to take no or more conservative action.

Overall, the new 3D model shows similar predictive performance and much tighter predictive bound when compared to the previous model. Cross-validation shows the new predictive model, constructed

from bin sizes likely reflective of SI dynamics, to accurately predict SI for data not used to develop the model. These improvements in prediction should translate to tighter GC, without compromising safety from hypoglycaemia.

8.5. Summary

SI plays a major role in any model-based GC protocol and SI forecasting is particularly important for managing dynamic ICU patients. It thus plays a leading role in the model-based STAR protocol. In particular, it enables a patient-specific approach to achieve better control and the use of forward stochastic prediction models enables safety and performance to be explicitly balanced in determining optimal insulin dosing.

This analysis has shown the positive impact of identifying prior change in a proven model-based SI metric on the prediction of likely future SI distribution ranges. A new 3D model was developed, achieving similar predictive power as the previous model, while significantly reducing the width of the 5^{th}-95^{th} percentile prediction range for more than 77% of the hours of data. This outcome ensures that over three-quarters of patient hours will be treated less over-conservatively. Equally, it also ensures the remaining quarter of patient hours are not treated aggressively (under-conservatively), and thus improves safety. Both outcomes will improve the performance, safety and patient specificity of GC, and thus patient outcomes.

The analysis in this chapter showed forward predictions were significantly improved when using the new 3D model. However, the methodology used here suffer from low resolution. More robust methods should thus be used to generalise these observations to a larger domain, and potentially, assess the impact of this new model on GC outcomes.

Chapter 9: Characterising Variability – Development and In-Silico Validation of a 3D kernel density stochastic model

The quality of GC resides in the ability to adapt treatment to time-varying patient-specific needs, which is a function of the level of difficulty of control [85, 98, 135]. This control difficulty mainly captured by SI variability [85, 135], where variability extremes can lead to hyper- or hypo- glycaemia for a given insulin and nutrition intervention, both associated with increased morbidity and mortality.

In the previous chapter, usage of prior temporal information of SI evolution (%ΔSI) demonstrated improved, more personalised accuracy in predicting future intra-patient SI variability [126]. However, the methodology had an important limitation in lacking model resolution and definition. In response, this chapter develops a new 3D stochastic model, using robust kernel density methods. It is then validated using virtual trials.

This chapter presents results published in [248], [249], and [127].

9.1. Introduction

Control difficulty is defined by SI variability [135], both physiologically and in model-based control, where variability extremes can lead to hyper- or hypo- glycaemia for a given insulin and nutrition intervention. The accurate prediction of future SI evolution is thus a key element for the quality of GC. The current stochastic models [124, 125] have been shown potentially over-conservative in the Chapter 8 due to large prediction bands [126]. Wide prediction bands can limit insulin dosing, resulting in lower insulin doses to avoid stochastically forecasted hypoglycaemic risk.

While encouraging, the method presented in Chapter 8 [126] lacks model resolution and definition, making comparison with the current 2D stochastic model used in STAR hard. This chapter thus aims to develop a new 3D stochastic model using a multivariate kernel density estimation method, similar to the one used for the current 2D stochastic model [125], accounting for prior knowledge of SI evolution. In contrast to the previous analysis [126], SI_{n-1} and SI_n are used to determine likely future SI_{n+1}, instead of SI_n and $\%\Delta SI_n$. However, $\%\Delta SI_n$ is still intrinsically captured by SI_{n-1} and SI_n, as a magnitude change instead of percentage change. Overall, the added input compared to the 2D stochastic model can provide higher patient-specificity, allowing more accurate insulin dosage for the patient, while the kernel density approach can provide smoother prediction intervals across the range of model inputs (SI_{n-1}, SI_n).

As already explained in Chapter 8, a wider future prediction range for SI would suggest higher potential variability, thus lower insulin rates will likely be recommended. In contrast, tighter prediction bands would suggest lower variability and thus, potentially higher insulin recommendation. In addition, this study assesses the impact of this new 3D stochastic model on GC performance, using validated virtual trial simulations [94, 128] (Chapter 3).

9.2. Methods

To account for patient-specific metabolic variability, and thus assess unexpected potential changes in metabolic response to insulin, [125] introduced a probabilistic model predicting likely future 1-3 hourly change in SI level (SI_{n+1}, SI_{n+2}, SI_{n+3}). These predictions are only based on current identified patient metabolic condition (SI_n). This stochastic model was built using a two-dimensional kernel density estimation method on population data, and led to the emergence of the first successful risk-based dosing

approach for GC [95, 96]. The kernel density estimation method enables high resolution behaviour estimation of a specific parameter based upon its prior evolution or state, even where specific data points may be scarce [142, 250, 251].

This study extends the originally used bi-variate kernel estimation method to tri-variate. The predictions of future SI_{n+1}, SI_{n+2}, and SI_{n+3} are thus determined using two inputs (SI_{n-1}, SI_n) to potentially increase patient-specific variability forecasting, which could also result in better overall safety and performance for STAR GC decision making. In particular, this choice of data triplets (SI_{n-1}, SI_n) $\rightarrow$ $SI_{n+1,2,3}$, adds patient specificity to the $SI_n \rightarrow SI_{n+1,2,3}$ 2D model by making these distributions a function of more prior states. This difference thus includes a greater part of the patient-specific evolution, and thus will further characterise patients, creating greater personalisation in the GC predictions based on thus enhanced stochastic model. It thus assumes there will be measurable differences in the predicted $SI_{n+1,2,3}$ distributions by this added data, compared to those from the 2D model. Importantly, the 3D approach significantly increases the data requirements for model generation, resulting in the use of a much larger data set size (~60000 hours) compared to previous studies [125], which is uniquely available as a result of regular use of STAR in multiple centres.

9.2.1. From Data Density to Condition Probability

SI in this study can be considered a second order finite Markov chain, where the current state depends only on its two prior states. Therefore, the conditional probability distribution of the future SI_{n+1} is a function of SI_n and SI_{n-1} states which can be expressed:

$$P(SI_{n+1} \mid SI_n, SI_{n-1}, ..., SI_0) = P(SI_{n+1} \mid SI_n, SI_{n-1}) = \frac{P(SI_{n+1}, SI_n, SI_{n-1})}{P(SI_n, SI_{n-1})} \qquad (9.1)$$

where the right-hand expression is derived from the general product rule. Kernel-density methods are used to estimate the joint probability $P(SI_{n+1}, SI_n, SI_{n-1})$ and $P(SI_n, SI_{n-1})$ using tri- and bi- variate Gaussian kernel density estimator functions [142]. Therefore, the conditional probability of SI_{n+1} taking a specific value can be calculated using the identified SI_n and SI_{n-1} values, such that:

$$\frac{P(SI_{n+1}=z \mid SI_n=y, SI_{(n-1)}=x)}{P(SI_n=y, SI_{n-1}=x)} = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{K_{hx_i}(u_{x_i})}{p_{x_i}} \frac{K_{hy_i}(u_{y_i})}{p_{y_i}} \frac{K_{hz_i}(u_{z_i})}{p_{z_i}}}{\frac{1}{N}\sum_{j=1}^{N} \frac{K_{hx_j}(u_{x_j})}{p_{x_j}} \frac{K_{hy_j}(u_{y_j})}{p_{y_j}}} \quad (9.2)$$

where $K_h(u)$ denotes the Gaussian kernel density function $K_h(u) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{1}{2}\left(\frac{u}{h}\right)^2}$, centred on u with variance h, constructed using the available N data points [250, 251]. To optimize the approximation of data behaviour, the variance h, or scale factor, is determined using the Silverman's general rule of thumb (ROT) [142, 250], weighted according to local data density and defined:

$$h = 0.9686\sigma\left(\frac{mR^3 N^{\frac{3}{7}}}{Z}\right)^{-\frac{1}{7}} \quad (9.3)$$

Where σ is the SD of the data, m is the number of data point within a radius $N^{-\frac{1}{7}}$ after orthonormalisation of the data [125], and R is the radius from the origin encompassing Z*N data points (0 ≤ Z ≤ 1).

This rule assumes data has an underlying normal distribution [250]. Non-negativity is ensured by normalizing each Gaussian function to the positive defined domain such that for each $(SI_n = y, SI_{n-1} = x)$ pair, there exists an estimated conditional probability function $P(SI_{n+1}=z \mid SI_n=y, SI_{n-1}=x)$ where $\int P(SI_{n+1}=z \mid SI_n=y, SI_{n-1}=x)dz = 1$ is satisfied [142]. Normalization is achieved by dividing each kernel density function $K_{x,y,z}(u)$ by the area under each gaussian curve between zero and infinity:

$$p_x = \int_0^\infty K_x(u)dx,\ p_y = \int_0^\infty K_y(u)dy,\ p_z = \int_0^\infty K_z(u)dz \quad (9.4)$$

This forces x, y, and z to be ≥ 0, thus ensuring physiological validity of SI values. An example of the resulting uni-, bi-, and tri- variate Gaussian kernel density estimation for 10 data triplets is shown in Figure 9.1.

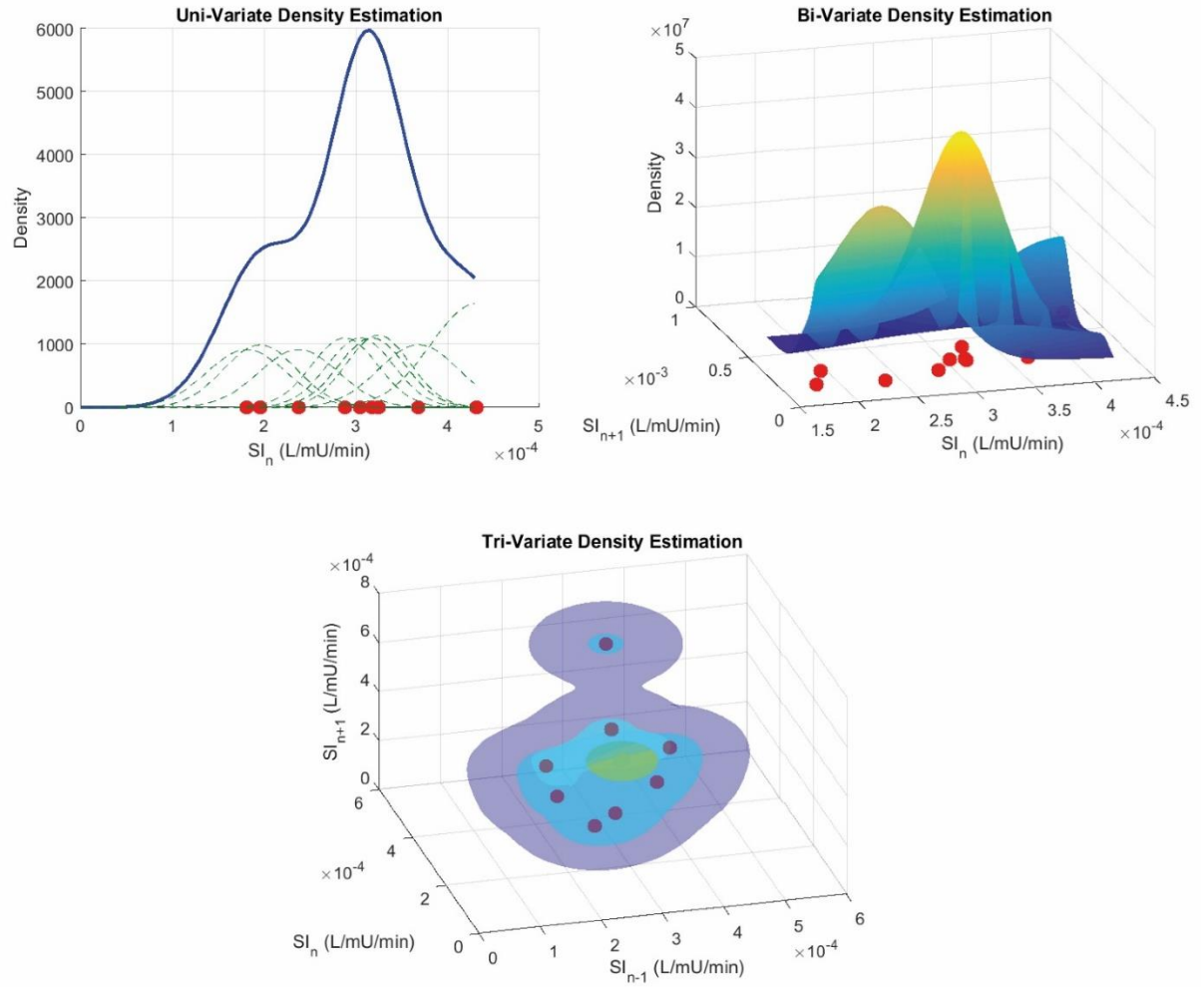

Figure 9.1 – Uni-, bi-, and tri- variate kernel density estimation for 10 data triplets. Dashed green lines show Gaussian distributions around each data point, where the standard deviation is a function of local data density.

9.2.2. Patients and Cohorts

The cohort used in this study is similar to the one used in the previous Chapter 8. In total 681 patient episodes ≥ 10 hours and with initial BG ≥ 7 mmol/L are considered, corresponding to 59439 hours of control. This data set is much larger than the compared to the one used for the 2D stochastic model [125], and ensures high data density for the method presented. SI is identified hourly for each patient using integral-based fitting method and a total of 58539, 57840, and 57141 data triplets (SI_{n-1}, SI_n, SI_{n+i}) for i = 1, 2, and 3 hours forward, respectively, are created.

9.2.3. Validation and Comparison Analysis

The 2D and 3D stochastic models are built and compared using five-fold cross-validation, where the resulting training (80%) and testing (20%) sets are believed to be statistically representative of the general dataset, minimizing bias and variance in the validation [252]. Patients are thus randomly divided

into 5 equally sized groups, models are built using 80% of patient episodes (4/5 groups), and the other 20% of patients (1/5 groups) are used for validation. As Silverman's ROT for multivariate kernel density estimation assumes data has a Gaussian distribution [250], and SI has a log-normal distribution, the logarithmic domain is chosen here to build the model.

The 25^{th}-75^{th} and 5^{th}-95^{th} percentile ranges are computed for both models. Tighter prediction ranges for future SI_{n+i} would suggest likely lower future variability. In this case, the future potential variation in SI being smaller, STAR can provide insulin with less risk and greater certainty, and thus potentially more aggressively (higher insulin rates) with equal safety. On the other hand, wider prediction bands would suggest higher future variability and, thus, more conservative dosing of insulin is necessary to avoid hypoglycaemia. Forward predictive power and model accuracy are compared using the percentage of accurate predictions within these two ranges. The expected accuracies are 50% and 90%, respectively, where greater conformation of an independent cohort to these expected outcomes, both overall and per-patient, would indicate the 3D methods more accurately capture SI dynamics to predict future SI.

Finally, to assess clinical impact, validated virtual trials on virtual patients are simulated to assess the new model's ability to control patients. Such virtual trials enable comparison of glycaemic outcomes from different GC designs, on the same underlying patients. In summary, virtual patients are characterized by their identified patient-specific SI traces generated from clinical data, and can be used to test a range of new protocols or technologies [91, 98]. They are well-validated in their independence from the data used to create them and their accuracy [94, 128], their ability to predict trial outcomes [96, 105, 129] and in clinical use to guide care in STAR [87, 95]. The underlying model is also well-validated in SI testing and similar clinical studies [130, 136, 137, 156]. These virtual trials have been validated in previous studies [94, 128], and are used here to simulate STAR using either the 2D (STAR-2D) or 3D (STAR-3D) stochastic model.

Unlike most GC protocols, STAR has the ability to modulate both insulin and nutrition inputs. Enteral nutrition can be lowered if the maximum allowed insulin is not sufficient to decrease BG levels, often occurring for very resistant patients with low SI and saturation of insulin dosing effects. In STAR, insulin is administered as boluses up to a maximum of 6U/hr, with an additional 3U/hr continuous infusion for highly resistant patients. Enteral nutrition administration can be modulated between 30-100% of the total

calorific GF if necessary. The original 100% GF for a patient is computed according to the standard 25 kcal/kg/day target [229] adapted based on age and sex. Further details are in [87, 134].

Safety and performance, administered insulin, and nutrition delivery are compared from these simulations. BG is resampled hourly, to allow fair comparison across the different measurement intervals. Safety is assessed by the %BG in mild (%BG ≤ 4.4 mmol/L) and severe (%BG ≤ 2.2 mmol/L) hypoglycaemia, and in hyperglycaemia (%BG > 8.0 mmol/L and %BG > 10.0 mmol/L). Performance is assessed by the %BG in the target band (4.4-8.0 mmol/L) and the median [IQR] BG levels achieved. Nutrition is reported as the percentage GF (%GF) achieved per-patient. In addition, workload is also compared, as the number of BG measurement per day, where a higher value indicates increased workload [174, 175].

9.3. Results

9.3.1. Forward Predictive Power Comparison Between the 2D and 3D Stochastic Models

A representation of the kernel density estimation is shown in Figure 9.2. The left panel shows the kernel density surface using the normal data, whereas the right panel shows the kernel density surface when data is transformed into the log-normal space to meet the normal distribution assumption under Silverman's ROT [250]. Clearly, log-normal data provides increased data density for higher SI ranges, where the raw data is sparser. Hence, this approach, taken for the first time here, potentially improves safety by better characterising SI potential variability for higher SI ranges, where the risk of experiencing hypoglycaemia due to insulin dosing is greater.

Cross-validation results of the forward predictive power for both models are summarised in Table 9.1. Additionally, the resulting 5^{th} and 95^{th} percentile predictions for each model are shown in Figure 9.3. Both the 2D and 3D models have close to 50% (~53% vs ~51%) and 90% (~91% vs. ~90%) predictions in the 25th-75th and 5th-95th percentile ranges respectively. However, the prediction ranges are generally narrower (~70% of hours) in the case of the 3D model. An example of the evolution of SI for a patient and the 2D and 3D predictions ranges for a specific virtual patient is shown in Figure 9.4. In addition, the median [IQR] percentage predictions in the 25th-75th and 5th-95th percentile prediction

ranges are closer to the expected 50% and 90% for the 3D model, suggesting the 2D model is too conservative for most patients.

To characterise the difference in prediction ranges from both models, the percentage change in the 5th-95th percentile range widths are computed for every prediction and the median [IQR] of percentage change is reported in Table 9.1. The high prediction performances are achieved with significantly 15.5-24.4% tighter 5th-95th percentile prediction range 69.9-73.8% of the time, and 14.8-22% wider otherwise. The median [IQR] 3D/2D prediction width ratios as a function of the hour-to-hour percentage change in of SI (%ΔSI) are shown in Figure 9.5, where clearly, prediction bands are typically tighter when %ΔSI is within ±20%. Overall, the new model thus better captures patient-specific differences from this more optimal model.

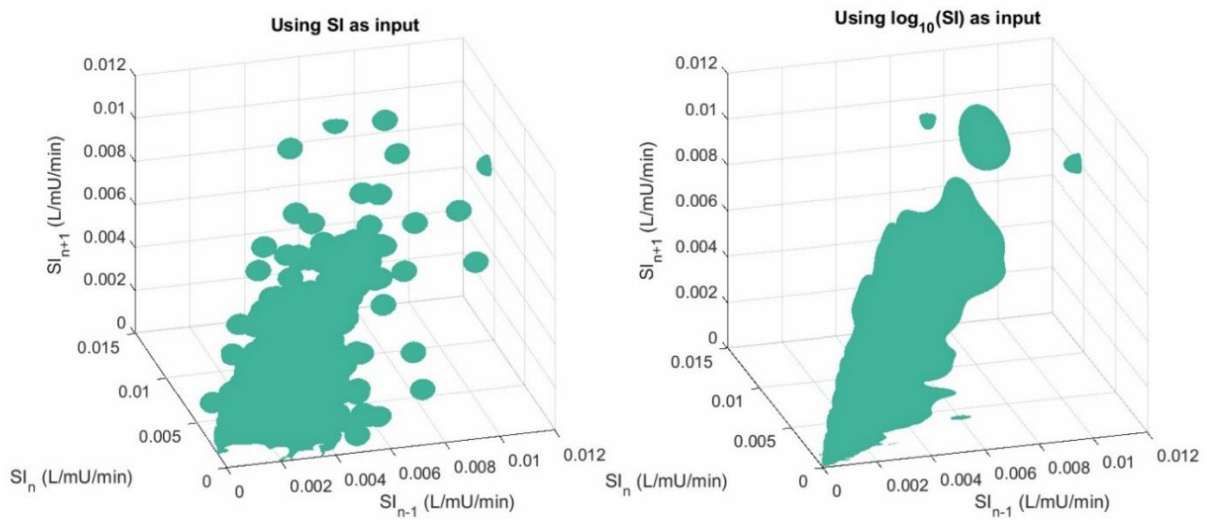

Figure 9.2 – Graphical representation of kernel density estimation using raw data (left) or logarithmic transformed data (right).

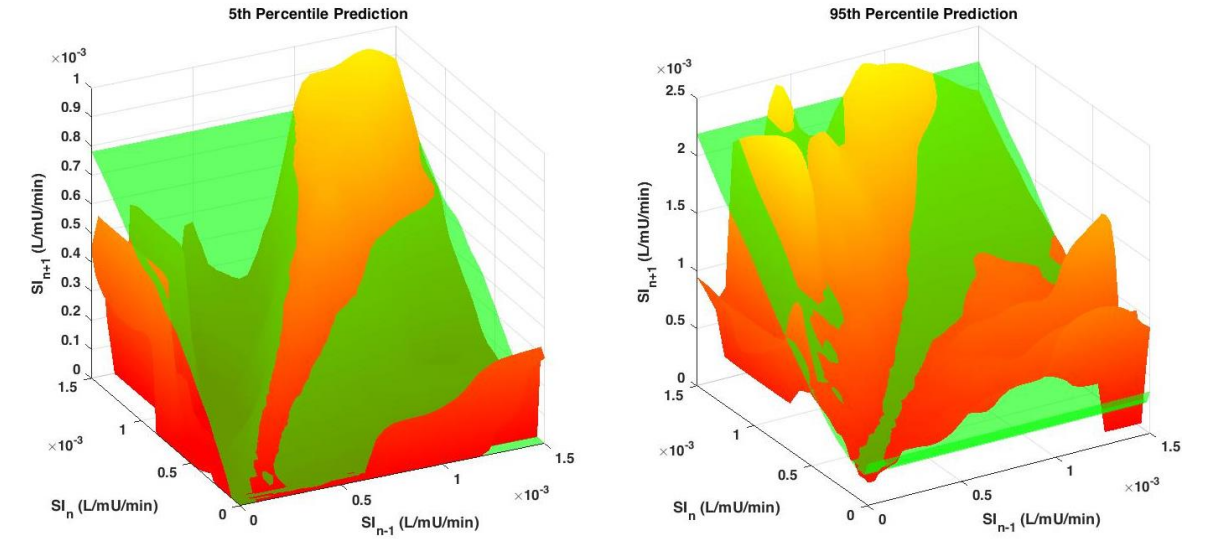

Figure 9.3 – Comparison between the 5th (left) and 95th (right) percentile predictions of likely future SI for the 2D (green) and the 3D (orange) models. The 2D model is constant across SI_{n-1} whereas the 3D model varies.

Table 9.1 – Five-fold cross-validation results summary of cohort forward predictive power and prediction range comparison between the 2D and 3D stochastic models.

		1-hourly	2-hourly	3-hourly
	Total predictions	58539	57840	57141
2D model	% predictions in 25^{th}-75^{th}	55.9	53.4	52.6
	% predictions in 5^{th}-95^{th}	91.4	91.0	91.0
3D model	% predictions in 25^{th}-75^{th}	52.6	51.3	51.0
	% predictions in 5^{th}-95^{th}	90.5	90.2	90.2
3D vs. 2D model	% of tighter predictions using 3D model	73.8	72.8	69.9
	% reduction in 5^{th}-95^{th} prediction width	24.4 [17.7 29.4]	17.9 [10.9 20.9]	15.5 [10.8 19.2]
	% of wider predictions using 3D model	26.2	27.2	30.1
	% increase in 5^{th}-95^{th} prediction width	22.0 [7.5 49.1]	16.4 [7.7 32.0]	14.8 [6.8 28.2]

Data is given as median [IQR] where appropriate.

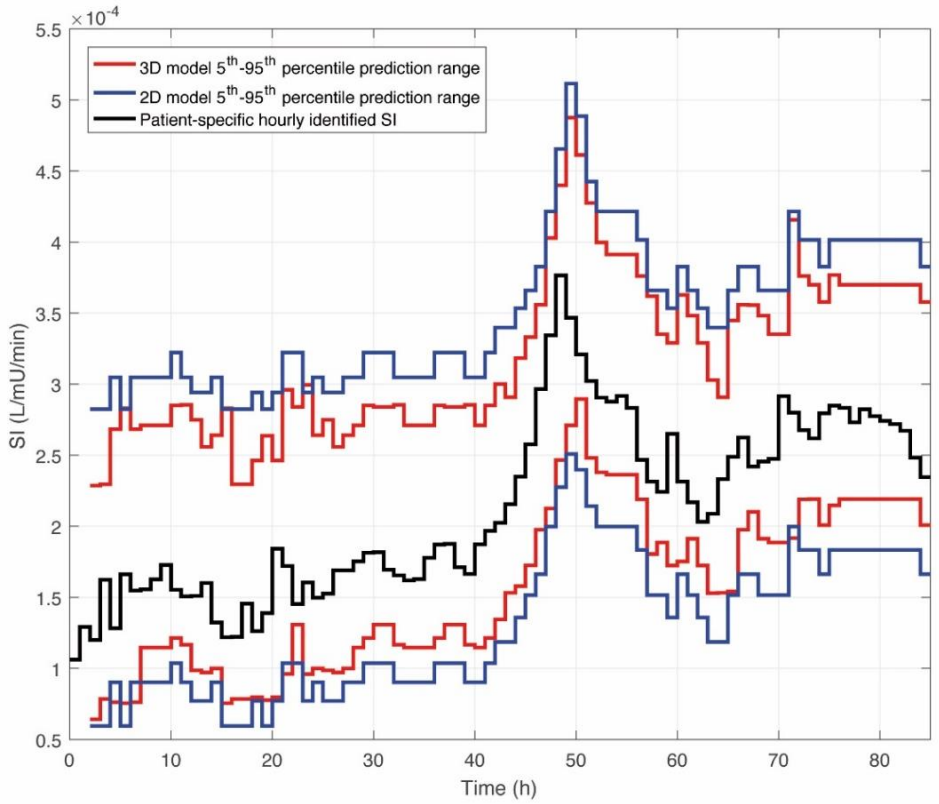

Figure 9.4 – Excerpt of SI evolution (black) and corresponding 2D (blue) and 3D (red) forward prediction ranges for a specific virtual patient. The 3D model prediction ranges are generally narrower.

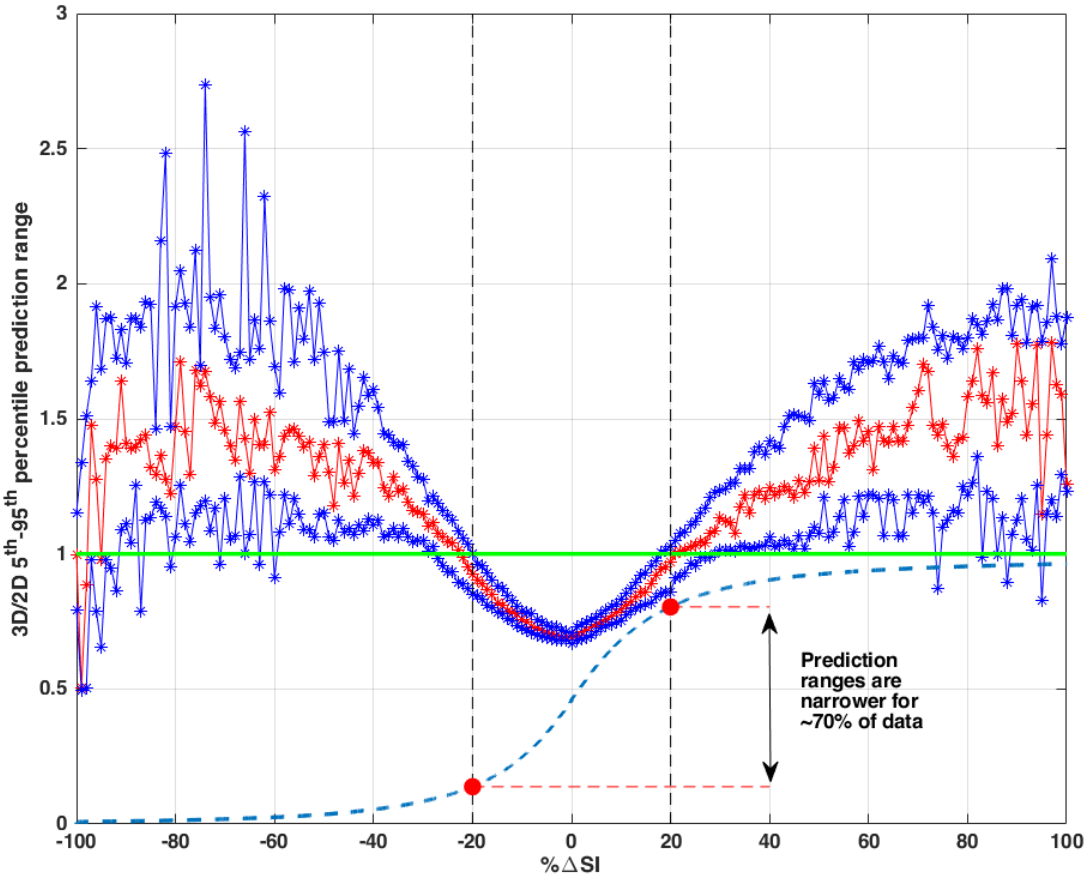

Figure 9.5 – Median [IQR] ratio between the 3D and 2D models 5th-95th percentile prediction width as a function of the hour-to-hour percentage change in SI (%ΔSI). The cumulative distribution function of %ΔSI is also shown in the blue dashed line.

9.3.2. Virtual Trials Results

Virtual trial results of STAR using the two different stochastic models are summarised in Table 9.2. Overall, both versions of STAR provided similar performance in terms of median BG [IQR] (6.3 [5.7, 7.0] vs. 6.2 [5.6, 6.9] mmol/L) and percentage time in the 4.4-8.0 mmol/L target band (88%). However, the overall %BG measurements shifted toward lower BG ranges using STAR-3D, with significantly higher %BG within 4.4-6.5 mmol/L and 4.4-7.0 mmol/L (61% vs. 56% and 75% vs. 72%, $p<0.01$ using χ^2 statistical test on proportions of measurements). In terms of safety, both models excel similarly with only ~2% BG < 4.4 mmol/L, ~1% BG < 4.0 mmol/L, and 0.03% BG < 2.2 mmol/L, despite STAR-3D administering higher median insulin (3.0 [1.5, 5.0] vs. 2.5 [1.5, 4.5] U/h). Slightly lower, but similar, %BG in 8-10 mmol/L (mild hyperglycaemia) for STAR-3D is also observed (7% vs. 8%). Finally, STAR-3D provided higher GF (97 [36, 100] vs. 95 [40, 100] %GF]).

Table 9.2 – Virtual trial results summary for STAR-2D and STAR-3D.

	STAR – 2D	STAR – 3D
Number of patients	681	681
Hours of control (h)	59073	59071
Total BG measurements	31248	31858
Workload (measurements per day)	12.7	12.9
Median [IQR] BG (mmol/L)	6.3 [5.7 7.0]	6.2 [5.6 6.9]
% BG in 4.4-6.5 mmol/L	56	61
% BG in 4.4-7.0 mmol/L	72	75
% BG in 4.4-8.0 mmol/L	88	88
% BG in 8.0-10.0 mmol/L	8	7
% BG > 10.0 mmol/L	3	3
% BG < 4.4 mmol/L	2.0	2.3
% BG < 4.0 mmol/L	0.9	1.0
% BG < 2.2 mmol/L	0.03	0.03
# patients < 2.2 mmol/L	11 (1.6%)	11 (1.6%)
Median [IQR] insulin rate (U/h)	2.5 [1.5 4.5]	3.0 [1.5 5.0]
Median [IQR] dextrose rate (%GF)	95 [40 100]	97 [36 100]

Data is given as median [IQR] where appropriate.

9.4. Discussion

The comparison between the 2D and 3D model clearly shows the new model's accuracy to predict future SI, with overall 15.5-24.4% tighter prediction range for more than 69.9-73.8% of the hours (Table 9.1). Typically, the prediction range is tighter when %ΔSI is within ±20% (Figure 9.5). On the contrary, the prediction range is wider when the variation is larger than ±20%. This key outcome thus suggests previous patient-specific metabolic variability has a direct impact on future SI forecasting.

More specifically, this 3D model shows stable patients, with low previous variation in SI, tend to remain stable, whereas more variable patients are more likely to have bigger future metabolic variations, clearly shown in Figure 9.3 and 7. Hence, the 2D stochastic model is over-conservative in terms of insulin intervention for most patients. The 3D approach allows STAR to select more aggressive insulin dosing more than 69.9% of the time, while ensuring safety, using the proven risk-based dosing approach. Therefore, the resulting greater patient-specificity implies better GC with lower glycaemic variability, and improved glycaemic outcomes.

The predictive power and tighter prediction ranges presented in this chapter are similar to those in Chapter 8 [126]. However, in Chapter 8, bins were used to balance data density, clearly impacting model resolution. Kernel density methods enable higher, continuous, resolution in this case. In addition, there was a clear lack of model definition where 10% of data were outside the stochastic model range with

the previous method, while this new stochastic model covers the global variable state space. In addition, while this method is more robust, it also better manages outliers and does not completely discard them. Transforming data into the logarithm space enables to naturally account less for these outliers, based on the local density, while still providing a smooth, realistic transition to these values. This new 3D stochastic model will thus be conservative in these unlikely events. Finally, this analysis also enables to extend the 3D model to 2- and 3- hourly forward prediction, where the Chapter 8 only presented a first approach based on 1-hourly forward prediction.

Virtual trial results comparing STAR using the 2D and the 3D stochastic models confirmed these observations showing higher percentage time in normo-glycaemic ranges, with 5% (absolute) more time spent in the 4.4-6.5 mmol/L range, for similar incidence of mild hypoglycaemia (BG < 4.4 mmol/L). Additionally, the 3D model resulted in more aggressive insulin dosing and higher feed rates for a similar intervention workload. Higher caloric intake is associated with improved outcomes [134, 143, 253, 254]. These outcomes confirm the 3D stochastic model, using prior information in SI variability, achieves effective control for all patients using more aggressive insulin dosing without compromising safety. Hence, STAR-3D offers a more patient-specific control, better accounting for either stable or very variable patients, potentially resulting in improved patient outcomes.

More importantly, the slightly lower median BG using STAR-3D (6.3 [5.7, 7.0] vs. 6.2 [5.6, 6.9]) was achieved with significantly higher time (61% vs. 56%, $p<0.01$ using χ^2 statistical test on proportions of measurements) in the 4.4-6.5 mmol/L band and in the 4.4-7.0 mmol/L band (75% vs. 72%, $p<0.01$ using χ^2 statistical test on proportions of measurements). While the low values for these p-values could be influenced by the large dataset size [111, 112], this difference is also clinically significant since larger values in these ranges have been associated with improved outcomes and higher odds of living [76, 77, 79].

Additionally, there was a consistent, high, 88 %BG in target band (4.4-8.0 mmol/L). High percentage time in these ranges have all been associated with improved clinical outcomes in multiple independent studies [76-79]. These results, together with the minimal cohort risk of hypoglycaemia (<2%) and severe hyperglycaemia (<0.03%), prove the STAR framework design to be adapted for GC in critical care, to

provide safe, effective control for all patients, and show GC to lower target ranges to be possible without compromising safety.

It is also important to note specific safety benefits of this new model are hard to highlight. First, because hypoglycaemia is extremely infrequent in STAR, unlike many other protocols failing to achieve safe control [46, 47, 49, 50, 52, 255]. Hence, the few hours where the 3D model enables a gain in reducing potentially very harmful hypoglycaemia due to highly variable SI are hard to see in the results, and overwhelmed in the overall high effectiveness of STAR. Thus, we examine improved performance more deeply, which is also beneficial for patients with equivalent safety.

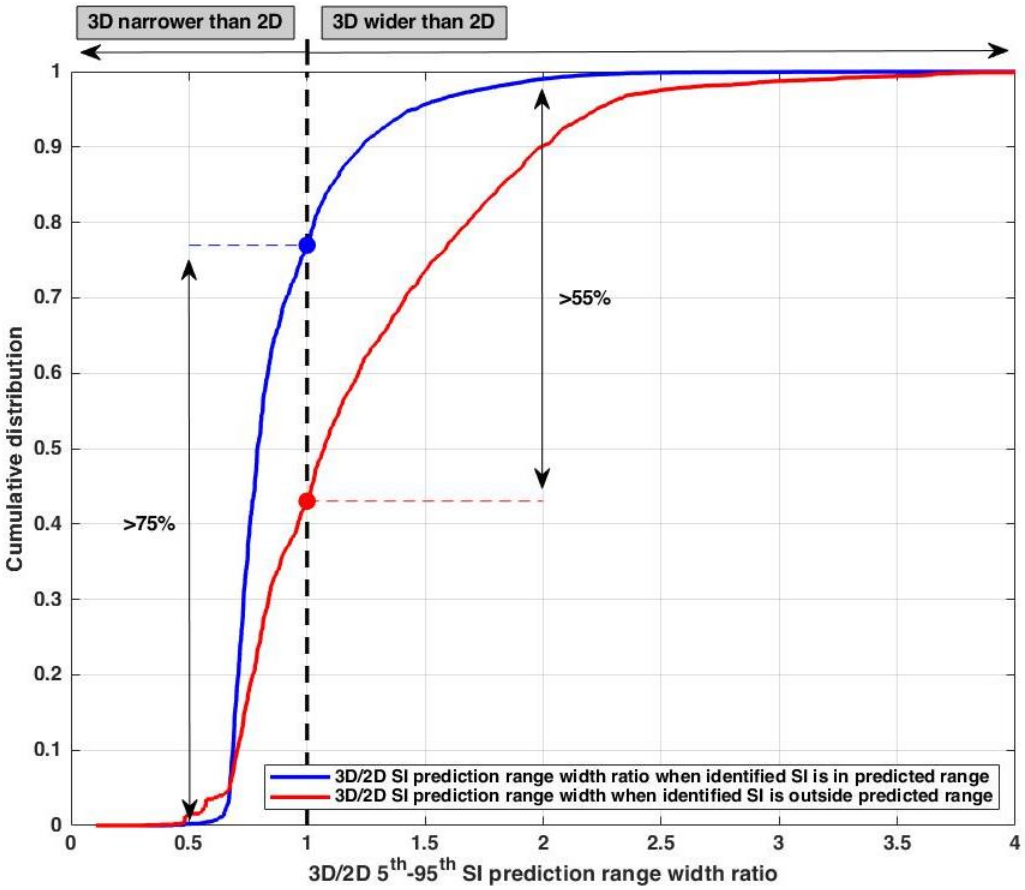

Figure 9.6 – Prediction range ratios CDFs when identified SI is within predicted ranges (blue) or outside (red).

To further illustrate this issue, the following CDFs of the ratio between the 5th-95th percentile range widths of each model when the subsequent SI value is within the predicted range and when the prediction is outside this range is shown in Figure 9.6. When SI is within the predicted range (~90% of hours, Table 9.1), the 3D model prediction band is tighter >75% of the time. However, when the subsequent SI value is outside the predicted range (~10% of the time), the 3D model is already >55% of the time wider than the 2D model. This result suggests when the subsequent SI value is outside the range, the 3D model is

generally more conservative (with a wider interval predicted) despite SI being outside predicted range. However, when the subsequent SI is within the predicted range, it is far narrower. Thus, the 3D model is overall safer.

While the difference in the two models shown in Table 9.1 is quite important, and the virtual trials showed higher performance (Table 9.2) at equal safety, a greater difference in glycaemic outcomes might have been expected. First, this difference shows how the STAR framework is consistent and manages to control patients in a safe manner. Second, the difference in SI prediction ranges between these models may not be big enough to change the discretized insulin interventions in STAR, as the controller is limited to 0.5 U/h increments. More specifically, in [135], an analysis suggests a change below 12-15% in SI levels can be considered clinically equivalent, limiting some impact on GC recommendations.

STAR treatment selection relies in putting the 5^{th} percentile of predicted BG outcome on the lower target band limit. Hence, it mainly uses the 95^{th} percentile of predicted SI. Looking deeper at the 95^{th} percentile difference between the 2D and 3D model, there is median reduction of ~6%. This difference may not be enough to significantly change the administration rate of insulin, leading to more similar glycaemic outcomes than would be obtained if insulin delivery rates were analogue to a smaller resolution, which would be less clinically feasible in workload and potential error [152].

As reflected in these results, using more information to better predict how likely patient-specific metabolic conditions will change seems a good approach to improve control in the STAR framework. More specifically, using more prior identified SI values also reduces the impact of direct measurement errors or identification errors for future prediction [170]. While one could think to extend this method to more dimensions, the danger would be to over fit the data and/or suffer from low data density, resulting in undesired behaviour for higher computational costs.

However, other parameters could be useful to improve both predictions and GC outcomes. In [256, 257], BG data is used as an entry with current SI level to forecast metabolic variability. In doing so, not only it potentially can improve control safety and efficacy, but it also allows to identify specific behaviour in the data, reflected by the resulting estimated distributions. In particular, [256, 257] observed typical underestimation of SI changes at lower BG values and vice-versa. Hence, more work could be done to

identify possible critical factors or parameters allowing to further improve prediction of important changes in metabolic variability and SI.

The bi-variate kernel density estimation method requires much fewer total data to create an effective model for use in clinical practice compared to the tri-variate model presented. However, the 3D stochastic model demonstrated better performance and equivalent safety in this study due to the much higher number of data triplets (~60,000 vs. ~20,000) available from the larger population data set used in this study than in creating the 2D stochastic model [124, 125]. In addition, the equivalence across the virtual trial five-fold cross validation results suggest the stochastic models were created on enough data to be robust and the data used was representative of a general ICU population.

The interpretation of these results has some limitations. Virtual trials represent realistic glycaemic outcomes in perfect implementation conditions, fully compliant to the protocol [152]. Glycaemic outcomes will likely differ at least somewhat in a real clinical environment. However, these virtual trials have been validated and shown to well capture the overall potential glycaemic outcomes [94, 128]. In addition, compliance to STAR is very high in regular clinical use [87, 258].

9.5. Summary

Tri-variate kernel density estimation methods are used here to build a new 3D stochastic model forecasting likely future changes in SI based on its prior 2 states. This 3D stochastic model shows similar, high, forward predictive power compared to the previous 2D version, but achieved with 15-25% tighter prediction ranges more than 70% of the time. This suggests the 3D stochastic model better predicts future SI dynamics and thus offers greater personalisation of care than the prior 2D model.

Virtual trials using this model showed similar GC safety and better performance based on higher time in the normoglycaemic intermediate ranges (4.4-6.5 mmol/L and 4.4-7.0 mmol/L), resulting in slightly lower median BG levels for similar workload. These improvements are due to greater personalisation of care, and were achieved by using higher insulin rates and slightly higher nutrition rates in cases where possible and as enabled by the tighter prediction ranges offered in over 70% of interventions. These results suggest the implementation of this new 3D stochastic model within the STAR framework could potentially improve patient clinical outcomes resulting from improved GC.

Chapter 10: Clinical Trial of STAR-3D Stochastic Model

Virtual trials results using the new 3D stochastic model in STAR have shown potential clinical benefits [126, 127]. More specifically, higher performance can be achieved without increasing the risk of hypoglycaemia or increasing workload, resulting from more accurate prediction of intra-patient variability. This new STAR-3D model was tested in a pilot trial as the clinical standard of care. This chapter presents and compares clinical trial results of STAR using the 3D stochastic model to prior STAR-2D results, as implemented in the Christchurch Hospital ICU, New Zealand.

10.1. Introduction

The previous chapters have shown personalised GC solutions in ICUs are essential for GC safety and efficacy. More specifically, directly quantifying and accounting for inter- and intra- patient variability has been identified as of primary importance for GC protocol designs. STAR is a model-based GC framework identifying patient-specific SI [130, 131] and forecasting future metabolic variability, enabling a unique risk-based dosing approach directly accounting for both inter- and intra- patient variability [95, 96, 123]. STAR is fully computerized, adjustable to different ICU practices, and has shown positive, near identical results in multiple centres, which no other GC protocol has done to date [87].

A 3D stochastic model was developed Chapter 8 and Chapter 9 to enable greater personalisation and precision [126, 127]. This new model added a second input parameter (SI_{n-1}) to the 2D stochastic model input (SI_n), to better predict future SI_{n+i} variability based on (SI_{n-1}, SI_n). Compared to the 2D stochastic model, prediction accuracy was modestly improved, but it also provided significantly tighter prediction ranges, enabling greater precision. Virtual trial results comparing STAR with both stochastic models showed similar safety and improved performance while providing more insulin and nutrition, compared to the 2D stochastic model [127]. Overall, these results justify the use of this model in clinical practice to improve patient care and outcomes.

This 3D stochastic model used in STAR (STAR-3D) has been implemented at Christchurch Hospital, New Zealand, to validate the results. The STAR-3D protocol is implemented as a new standard of care alongside a technology upgrade, due to clinical confidence in the STAR protocol, for a pilot trial. The clinical results to date are analysed and compared to previous published results of STAR using the 2D stochastic model (STAR-2D) in the same ICU [87].

10.2. Methods

10.2.1. STAR-3D Glycaemic Control Framework

The STAR framework has been developed in Chapter 3. In its original version, STAR predicts potential future metabolic variability based on the identified, patient-specific, current SI_n level, using a 2D stochastic model [124, 125]. The 3D stochastic model now uses also SI_{n-1} providing additional temporal information for more accurate predictions of SI_{n+1} [126, 127]. One of the main advantages of this new

stochastic model is it better characterises intra-patient variability, and, thus, better captures the behaviour of this variability and adapts predictions accordingly. In essence, greater personalisation offers the opportunity for greater precision in care.

The new 3D stochastic model showed stable patients tend to remain stable, with less future potential variability, while more variable patients are subjected to higher potential variability. STAR uses the 5^{th}-95^{th} prediction of future evolution of SI, to adapt insulin and treatment so the corresponding 5^{th}-95^{th} prediction of BG levels overlaps the clinically set target band [95, 96, 123]. Thus, more accurate prediction of this SI range can directly improve and add precision to STAR GC control outcomes. The full methodology and advantages of this new approach are presented in Chapter 3.

STAR is fully computerised and has been adapted to use the 3D stochastic model. This process did not require ani significant changes to the original STAR protocol, minimising/mitigating potential software errors. This software change enables clinical testing and use as a standard of care in Christchurch, New Zealand, which already employs STAR (2D) as a standard of care.

The starting criterion for this trial is 2 consecutive BG measurements > 8.0 mmol/L or clinical choice. BG assays are made using standard glucometers (Accu-Check® Inform II, Roche, Switzerland). Insulin is administered as boluses through an intravenous catheter, with increments of a maximum +2 U/h between consecutive measurements, and limited to a maximum size of 6U/h. For very insulin-resistant patients, an additional background insulin infusion rate of 3 U/h can be administered.

In STAR, both insulin and nutrition are modulated. Enteral nutrition can thus be adjusted treatment to treatment by maximum 30% of the original 100% GF rate, going no lower than 30% of the original GF. The original GF rate is derived from a daily base rate of 25 kcal/kg/day nutritional intake [229], and adapted for STAR GC based on age, sex, and frame size/weight [134]. The target band is 4.4-8.0 mmol/L, and STAR is stopped if the patient BG is stabilised for 6 hours (BG in target band and insulin ≤ 2 U/h). Nurses are free to choose between the suggested 1-3 hour interval treatments, and to adapt rates according to their clinical judgment, if desired. The protocol is implemented on Android operated system tablets at the patient bedside. All of these criteria and approach are unchanged from using STAR-2D.

STAR is currently the standard of care in the Christchurch Hospital ICU, New Zealand. Hence, implementing this new version was a simple change of practice and did not require additional approval from the local ethics committee for the change or to audit results, based on prior approval for STAR-2D from the New Zealand Upper South Island Regional Ethics Committee.

10.2.2. Patients and Cohort

This analysis uses clinical data from 181 patients and 273 GC episodes included in STAR using the 3D stochastic model (STAR-3D) between April 2019 and January 2020. From these 273 episodes, 48 (18%) episodes were longer than 10 hours, 21 (8%) had average nutrition rates higher than 120% GF, and 4 (1%) targeting a range other than the clinically specified 4.4-8.0 mmol/L (Figure 10.1), leaning to 200 (73%) of episodes remaining for the analysis. These criteria cover the normal , per protocol use of STAR, and ensure a fair comparison with retrospective clinical data of 264 patients under the original version of STAR (STAR-2D), presented in Chapter 4, with the same conditions. Demographics of the resulting patient cohort is presented in Table 10.1, where STAR-3D and STAR-3D patients are similar in age, sex, severity of injury, and operative patients.

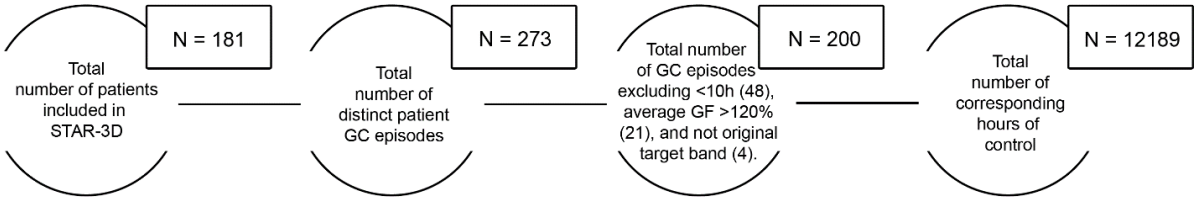

Figure 10.1 – GC episode selection from the original 181 patients (200 episodes) included in the STAR-3D clinical trial.

Table 10.1 – Demographics summary of patients included in the STAR-3D and STAR-2D clinical trials.

	STAR-2D	STAR-3D	P values
# episodes	330	200	/
# patients	264	146	/
# control hours	22372	12189	/
Percent male	66	71	0.27[a]
Age (years)	65 [55, 73]	65 [52, 72]	0.50[b]
APACHE 2	21 [16 25]	20 [15 26]	0.98[b]
APACHE 2 RoD	33 [15 51]	/	/
APACHE 3	73 [56 94]	75 [56 103]	0.23[b]
APACHE 3 RoD	26 [11 50]	23 [9 57]	0.75[b]
% Operative	43	39	0.47[a]

Data is given as median [IQR] where appropriate. Statistical testing using [a]Fisher exact test, or [b]Wilcoxon rank sum test.

10.2.3. Clinical Results Comparison Analysis

Clinical trial results of STAR-3D and STAR-2D results are compared. Safety, efficacy, BG achieved, insulin and nutrition rates, and workload are compared. BG is resampled hourly to allow fair comparison between protocols [87, 122]. Safety is compared using the percentage BG outside the target band (%BG < 4.4 mmol/L and %BG > 8.0 mmol/L), as well as the percentage BG and the number of patients below the severe hypoglycaemic threshold (%BG < 2.2mmol/L). Performance is analysed using the percentage BG in the 4.4-8.0 mmol/L target band and median [IQR] BG levels achieved.

Per-episode insulin (U/h) and nutrition rates (%GF) are also compared, and workload is assessed using average measurements per day. Additionally, the percentage BG in 4.4-6.5 mmol/L and 4.4-7.0 mmol/L are compared for each protocol to indicate tightness to lower-normal range. High percentage time in these bands are associated with improved outcomes in ICU patients [38, 76, 77, 79]. Results are examined at both cohort and per-patient/episode levels.

Due to the relatively large data sample sizes, bootstrapping methods are used to determine if the null hypothesis of samples being drawn from distributions of equal medians can be rejected or not, at a statistical significance of $\alpha=0.05$ (Chapter 2). The 95% CI of difference in bootstrapped medians is computed. Distributions are considered significantly different ($p<0.05$) if the 95%CI does not include the null hypothesis. The Fisher Exact test is used for proportion data, at the same significance level ($\alpha=0.05$).

As per protocol design and virtual trial results presented in Chapter 9, similar time in target band is expected, with a shift in BG levels toward lower ranges (4.4-6.5 mmol/L and 4.4-7.0 mmol/L). Additionally, higher nutrition and insulin rates are expected, with similar workload if starting glucose is similar. The main outcome of the study is to evaluate the impact on overall GC outcome using STAR-3D compared to STAR-2D.

Importantly, if one cohort has significantly higher starting BG than the other, the performance will be affected as it will typically take more time to reach the target. In addition, during this longer time, nutrition rates are potentially further reduced, also source of bias in the comparison. To avoid bias comparing on clinical results for cohorts with different starting BG, a secondary analysis calculates safety and

performance once the target band is reached. This analysis fairly captures each protocols' ability to safely and effectively control BG.

10.2.4. Virtual Trial Comparison Analysis

Another main advantage of virtual trials in the context of this thesis, is the ability to directly create virtual patients from clinical patients, and simulate the protocol to see the difference with clinical GC outcomes under full compliance and ideal conditions. Such simulations are possible because model-based SI is treatment independent (Chapter 3) [91, 94, 128]. Therefore, patient-specific SI evolution characterises patient metabolic evolution, regardless of the treatment received (ie: the identified SI trace evolution for a patient would have been identical under any other GC protocol).

SI is thus identified hourly [130, 131] from each clinical patient data to create virtual patients. Virtual trials using the STAR-3D (STAR-3D-VT) protocol are simulated, and results compared to STAR-3D clinical GC outcome. Large differences in glucose results would indicate reduced protocol compliance. Similar statistics and hypothesis testing as for the clinical results analysis are used to compare results.

10.2.5. Compliance Analysis

Nurse compliance to protocol is also analysed. Analysing compliance to protocol enables to identify and assess the potential impact of clinical staff deviations from original recommendations on GC outcomes. It also allows to determine whether protocol design is feasible. Hence, maximum treatment intervals suggested by STAR for each treatment is computed, as well as the number of time nurses picked the longest available treatment interval. Additionally, the percentage time clinical staff changed overrode original recommendation is recorded. Changes are considered as deviations from original protocol if they are made within 15 minutes after treatment selection and confirmation.

10.3. Results

10.3.1. Clinical Results

Cohort clinical GC outcome results for both cohorts are presented in Table 10.2. Performance is high in both cohorts, but STAR-3D has slightly lower time in the 4.4-8.0 mmol/L target band (78%) compared to STAR-2D (83%). Additionally, time in lower intermediate bands is also slightly lower for STAR-3D

(43% in 4.4-6.5 mmol/L and 60% in 4.4-7.0 mmol/L) compared to STAR-2D (44% and 62%, respectively). Finally, cohort median BG levels achieved in STAR-3D are slightly higher (6.7 [6.0 7.7] mmol/L) than STAR-2D (6.6 [6.0 7.4] mmol/L), but clinically similar (Chapter 2). The CDFs of resampled BG levels are shown in Figure 10.2, where the main difference is greater incidence of higher BG levels. Importantly, starting BG is higher under STAR-3D (10.5 [8.8 12.8] mmol/L) compared to STAR-2D (9.7 [8.4 11.5] mmol/L). Hyperglycaemia is higher in STAR-3D (8%) than STAR-2D (4%). However, hypoglycaemia was reduced with less %BG below target band on STAR-3D (0.9%) compared to STAR-2D (1.4%), and no patients with BG < 2.2 mmol/L (0%) versus 3 patients for STAR-2D (1%)

These cohort outcomes were achieved with slightly workload for STAR-3D compared to STAR-2D (13.8 vs. 12.9 measurements per day), significantly higher insulin rates (4.0 [2.0 6.0] U/h vs. 2.5 [1.0 4.5]U/h, $p<0.05$), and significantly higher nutrition rates (98 [80 100] %GF vs. 92 [71 100] %GF, $p<0.05$). At a cohort perspective, both protocols achieved similar high quality of control, with STAR-3D being safer regarding hypoglycaemic risk, and providing more nutrition than STAR-2D. However, STAR-3D required 1 extra measurement per day, perhaps due to the longer time to reach the target range due to higher starting BG.

Table 10.2 – Cohort clinical results summary for STAR-3D and STAR-2D.

	STAR-2D	**STAR-3D**	**P-value**
Number of episodes	330	200	/
Hours of control (h)	22372	12189	/
Starting BG (mmol/L)	9.7 [8.4 11.5]	10.5 [8.8 12.8]	**<0.05**
Total BG measurements	12030	6980	/
Workload (measurements per day)	12.9	13.7	/
BG (mmol/L)	6.6 [6.0 7.4]	6.7 [6.0 7.7]	**<0.05***
% BG in 4.4-6.5 mmol/L (%)	44	43	/
% BG in 4.4-7.0 mmol/L (%)	62	60	/
% BG in 4.4-8.0 mmol/L (%)	83	78	/
% BG in 8.0-10.0 mmol/L (%)	11	13	/
% BG > 10.0 mmol/L (%)	4	8	/
% BG < 4.4 mmol/L (%)	1.4	0.9	/
% BG < 4.0 mmol/L (%)	0.5	0.3	/
% BG < 2.2 mmol/L (%)	0	0	/
Insulin rates (U/h)	2.5 [1.0 4.5]	4.0 [2.0 6.0]	**<0.05**
Total hours not fed (%)	10	18	/
Dextrose rates excluding not fed hours (%GF)	92 [71 100]	98 [80 100]	**<0.05**

Data is given as median [IQR] where appropriate. Statistical testing using 95% CI bootstrapped BG medians difference. (*) indicates 95% CI bootstrapped BG medians difference is clinically equivalent.

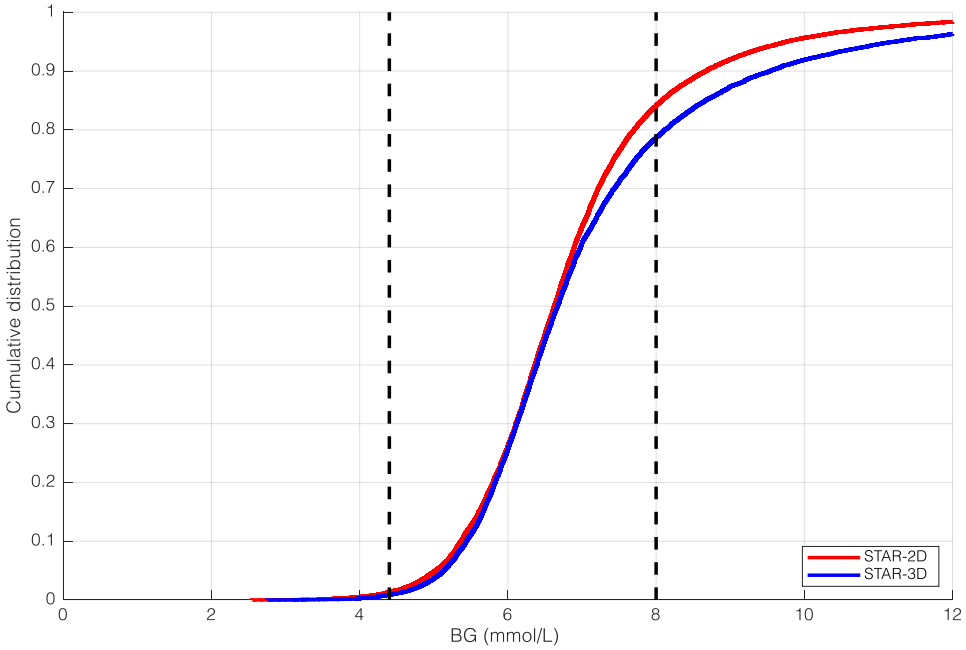

Figure 10.2 – Resampled BG cumulative distribution functions for STAR-2D (red) and STAR-3D (blue).

Table 10.3 presents per-episode clinical GC outcome results for both cohorts. Overall, high performance and safety was achieved, per-episode, similarly for both cohort (p>0.05). Only severe hyperglycaemia is significantly higher in STAR-3D compared to STAR-2D (4 [0 14] vs. 2 [0 6] %BG>10.0 mmol/L, p<0.05). Similar median BG are achieved in both cohorts. Per-episode 5th-95th, 25th-75th, and median BG achieved are presented in Figure 10.3, clearly showing similar median (Table 10.3), but clearly higher 95th percentile BG in STAR-3D.

Table 10.3 – Per-episode clinical results summary for STAR-3D and STAR-2D.

	STAR-2D	**STAR-3D**	**P-value**
Number of episodes	330	200	/
Episode length (days)	1.8 [0.9 3.6]	1.7 [0.8 3.5]	NS
BG measure/day per patient	13.6 [11.5 16.2]	13.9 [11.6 16.7]	NS
Starting BG (mmol/L)	9.7 [8.4 11.5]	10.5 [8.8 12.8]	**<0.05[a]**
Median BG (mmol/L)	6.5 [6.1 7.0]	6.6 [6.2 7.2]	NS
% BG in 4.4-6.5 mmol/L (%)	46 [27 61]	45 [29 61]	NS
% BG in 4.4-7.0 mmol/L (%)	65 [46 75]	63 [41 78]	NS
% BG in 4.4-8.0 mmol/L (%)	85 [73 93]	81 [65 92]	**<0.05[a]**
% BG in 8.0-10.0 mmol/L (%)	9 [4 17]	10 [5 18]	NS
% BG > 10.0 mmol/L (%)	2 [0 6]	4 [0 14]	**<0.05[a]**
% BG < 4.4 mmol/L (%)	0.0 [0.0 1.7]	0.0 [0.0 0.0]	NS
% BG < 4.0 mmol/L (%)	0.0 [0.0 0.0]	0.0 [0.0 0.0]	NS
% BG < 2.2 mmol/L (%)	0.0 [0.0 0.0]	0.0 [0.0 0.0]	NS
# patients < 2.2 mmol/L (%)	3	0	0.25[b]
Insulin rate (U/h)	2.5 [1.0 3.5]	3.5 [2.5 5.0]	**<0.05[a]**
# episodes with nutrition (%)	81	75	0.12[b]
Dextrose rate excluding not fed (%GF)	88 [63 100]	97 [80 100]	**<0.05[a]**

Data is given as median [IQR] where appropriate. Statistical testing using ([a]) 95% CI bootstrapped BG medians difference, or ([b])Fisher Exact test. NS = not statistically significant.

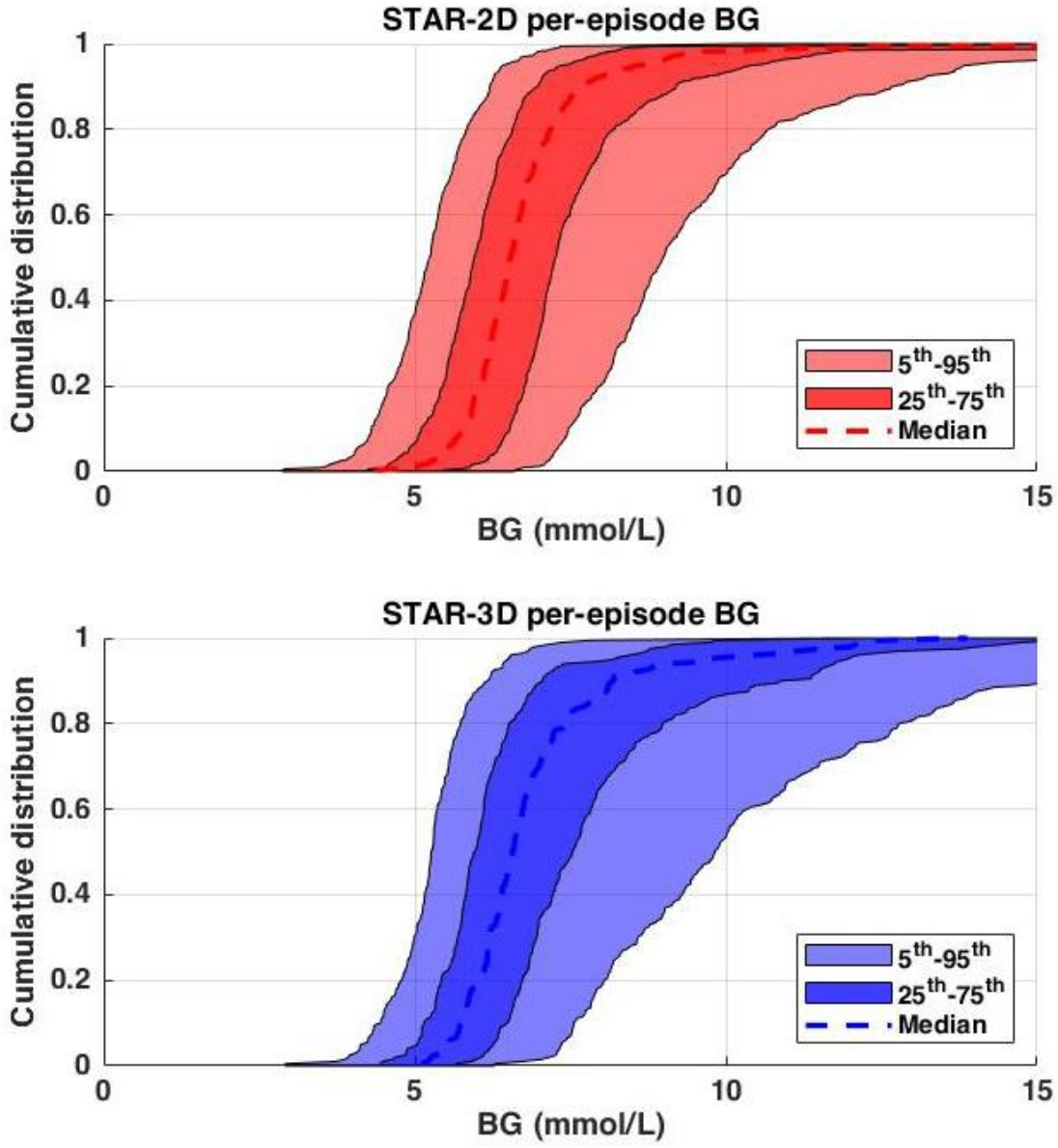

Figure 10.3 - Per-episode median BG as well as the 25th-75th and 5th-95th percentile ranges for STAR-2D (top) and STAR-3D (bottom).

Initial BG levels are significantly higher for STAR-3D than STAR-2D (10.5 [8.8 12.8] vs. 9.7 [8.4 11.5] mmol/L, $p<0.05$), as shown in Figure 10.4 (top panel). For example, a starting BG of 10.0 mmol/L is the ~40th percentile for STAR-3D in Figure 10.4, but ~60th percentile for STAR 2D. Thus, more patients start at relatively higher BG levels, impacting performance and workload. The bottom panel in Figure 10.4 shows STAR-3D (blue) provided significantly improved control ($p<0.05$) relative to higher initial BG (shift to the left) compared to STAR-2D (red). For example, the likelihood of a 25% reduction (BG/Initial BG = 0.75%) in BG levels is 50% for STAR-2D, and 63% for STAR-3D, a significant relative difference.

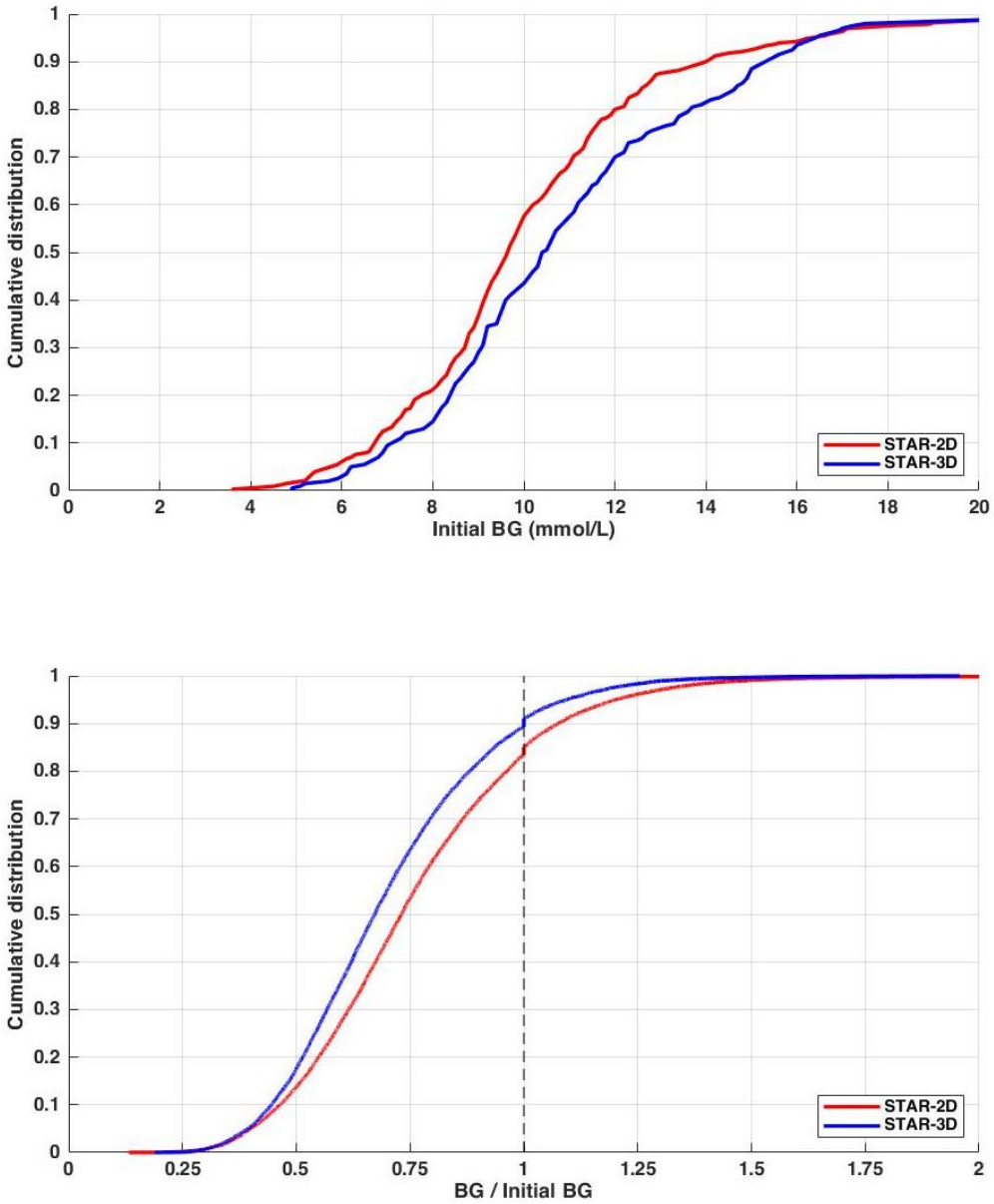

Figure 10.4 – Per-episode initial BG (top) and cohort BG / initial BG ratio (bottom).

Importantly, median [IQR] episode length is similar between STAR-3D and STAR-2D (1.7 [0.8 3.5] vs. 1.8 [0.9 3.6] days, $p>0.05$), as well as the median [IQR] workload (13.9 [11.6 16.7] vs. 13.6 [11.5 16.2] measures per day). Finally, STAR-3D provides significantly higher insulin (3.5 [2.5 5.0] vs. 2.5 [1.0 3.5] U/h, $p<0.05$) and nutrition (97 [80 100] vs. 88 [63 100] %GF, $p<0.05$) rates than STAR-2D, showing the potential for more aggressive insulin dosing enabled by the 3D stochastic model.

Thus, per-episode, STAR-3D achieved similar safety and efficacy GC outcomes, with higher insulin and nutrition rates, and higher starting BG. The significantly higher initial BG in STAR-3D most likely explains the higher %BG above target, the slightly reduced time in band, and the slightly higher workload, but these points are further analysed in Section 10.3.2.

10.3.2. Clinical Results Once Target Band Reached

Cohort and per-episode clinical GC outcome results for both cohorts are presented in Table 10.4 and Table 10.5. From all episodes, 328 (99.4%) for STAR-2D and 199 (99.5%) for STAR-3D reach the target band. STAR-3D episodes needed more time to reach the target (4 [2 7] h) compared to STAR-2D (3 [2 5] h), representative of higher starting BG in Table 10.4. At a cohort perspective, BG levels achieved are still clinically equivalent. Most importantly, the overall performance of STAR-3D is now much closer to STAR-2D than previously. Severe hyperglycaemia in STAR-3D (3%) is now also closer to STAR-2D (2%), and much lower compared to raw clinical results (8% and 4%, respectively). Incidence of hypoglycaemia did not change. Insulin and nutrition rates are still significantly higher in STAR-3D,

At a per-episode level, median BG remains clinically equivalent, and workload, despite not being statistically different (13.3 [11.0 16.0] vs. 13.1 [11.1 15.2] measures per day for STAR-3D and STAR-2D, respectively), is decreased compared to clinical raw data (13.9 [11.6 16.7] vs. 13.6 [11.5 16.2]), explained by discarding hours before entering BG target. Performance achieved by both protocols is now much closer (Figure 10.5), with no statistical difference in any intermediate ranges. In addition, per-episode median percentage BG in severe hyperglycaemia is also no longer significant, with minimal incidence. The only significant difference remaining, at a per-episode level, is the higher insulin and nutrition rates achieved.

Table 10.4 – Cohort clinical results summary for STAR-3D and STAR-2D once in target band.

	STAR-2D	**STAR-3D**	**P-value**
Number of episodes	328 (99.4%)	199 (99.5)	/
Starting BG (mmol/L)	9.7 [8.4 11.5]	10.5 [8.8 12.8]	**<0.05**
Time to target (h)	3 [2 5]	4 [2 7]	>0.05
Hours of control (h)	21484	11298	/
BG (mmol/L)	6.6 [5.9 7.3]	6.6 [5.9 7.4]	**<0.05***
% BG in 4.4-6.5 mmol/L (%)	46	47	/
% BG in 4.4-7.0 mmol/L (%)	65	65	/
% BG in 4.4-8.0 mmol/L (%)	87	85	/
% BG in 8.0-10.0 mmol/L (%)	10	11	/
% BG > 10.0 mmol/L (%)	2	3	/
% BG < 4.4 mmol/L (%)	1.4	1.0	/
% BG < 4.0 mmol/L (%)	0.6	0.3	/
% BG < 2.2 mmol/L (%)	0	0	/
Insulin rates (U/h)	2.5 [1.0 4.0]	3.5 [1.5 6.0]	**<0.05**
Total hours not fed (%)	13	22	/
Dextrose rates excluding not fed hours (%GF)	92 [74 100]	99 [80 100]	**<0.05**

Data is given as median [IQR] where appropriate. Statistical testing using 95% CI bootstrapped BG medians difference. (*) indicates 95% CI bootstrapped BG medians difference is clinically equivalent.

Table 10.5 – Per-episode clinical results summary for STAR-3D and STAR-2D once in target band.

	STAR-2D	STAR-3D	P-value
Number of episodes	228	199	/
Episode length (days)	1.7 [0.8 3.4]	1.5 [0.6 3.1]	NS
BG measure/day per patient	13.1 [11.1 15.2]	13.3 [11.0 16.0]	NS
Median BG (mmol/L)	6.5 [6.1 6.9]	6.4 [6.0 6.9]	NS
% BG in 4.4-6.5 mmol/L (%)	50 [31 65]	52 [35 70]	NS
% BG in 4.4-7.0 mmol/L (%)	71 [54 82]	71 [52 86]	NS
% BG in 4.4-8.0 mmol/L (%)	92 [82 100]	95 [82 100]	NS
% BG in 8.0-10.0 mmol/L (%)	4 [0 12]	3 [0 14]	NS
% BG > 10.0 mmol/L (%)	0 [0 0]	0 [0 0]	NS
% BG < 4.4 mmol/L (%)	0 [0 1.9]	0 [0 0]	NS
% BG < 4.0 mmol/L (%)	0 [0 0]	0 [0 0]	NS
% BG < 2.2 mmol/L (%)	0 [0 0]	0 [0 0]	NS
# patients < 2.2 mmol/L (%)	0	0	1.0[b]
Insulin rate (U/h)	2.0 [1.0 3.0]	2.5 [1.5 4.0]	NS
# episodes with nutrition (%)	73	74	0.23[b]
Dextrose rate excluding not fed (%GF)	88 [65 100]	95 [75 100]	**<0.05[a]**

Data is given as median [IQR] where appropriate. Statistical testing using ([a]) 95% CI bootstrapped BG medians difference, or ([b])Fisher Exact test. NS = not statistically significant.

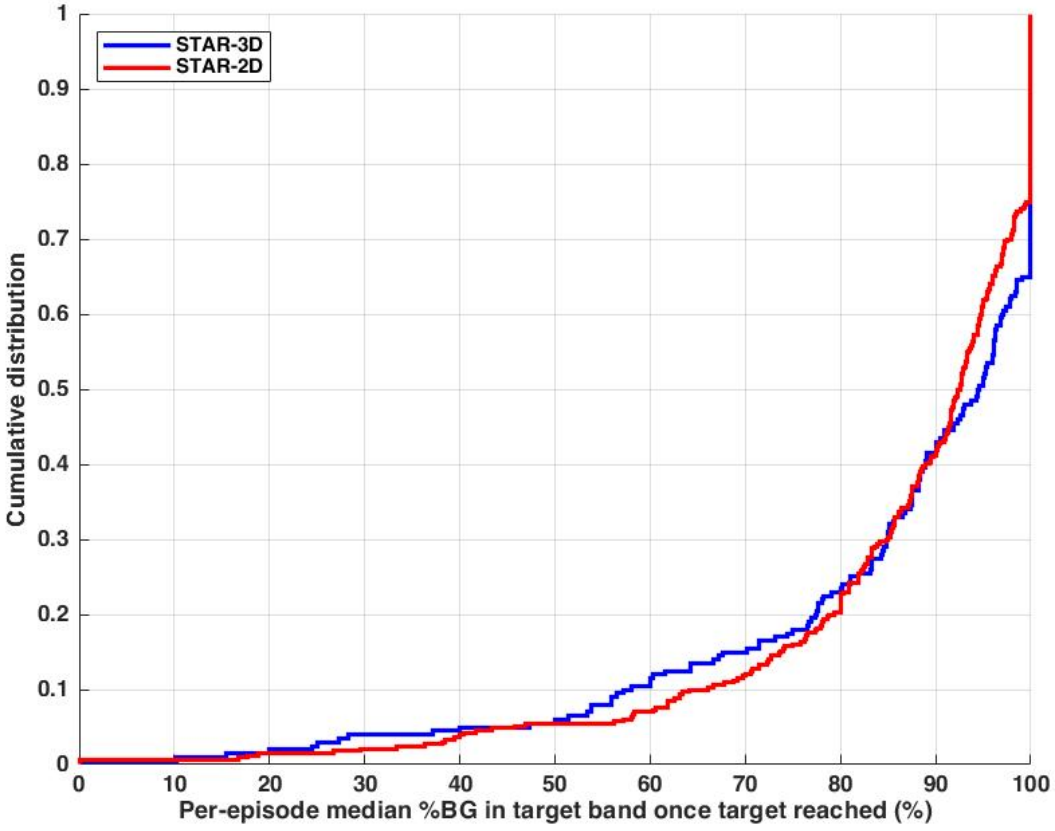

Figure 10.5 – CDFs comparison of the per-episode median %BG in target band for both protocols.

10.3.3. Virtual Trials Results

To compare and assess how STAR-3D would have performed in ideal conditions (full compliance to protocol), virtual trial of STAR-3D (STAR-3D-VT) are simulated on virtual patients, created based on clinical data from 200 GC episodes of the STAR-3D clinical trial. STAR-3D-VT reflect perfect compliance and minimal workload by selecting the greatest measurement interval offered. The results are compared to clinical results of STAR-3D in Table 10.6 and Table 10.7. Virtual trial overall cohort results of STAR-3D-VT show higher performance and similar safety than the clinical results of STAR-3D (Table 10.6). More specifically, while the %BG in target bands are similar (82% vs. 78% for STAR-3D-VT and STAR-3D, respectively), there is a clear shift to lower ranges with 52 %BG in the 4.4-6.5 mmol/L range achieved in virtual trials compared to 43% clinically. Lower median BG levels are thus achieved in virtual trials (6.4 [5.8 7.4] mmol/L) compared to clinical results (6.7 [6.0 7.7] mmol/L), but this difference is clinically equivalent. Insulin is significantly lower ($p<0.05$) clinically (4.0 [2.0 6.0] U/h) compared to simulations (5.0 [3.0 6.0] U/h), while nutrition is higher (100 [70 100] %GF for STAR-3D-VT vs. 98 [80 100] for STAR-3D, $p<0.05$).

Table 10.6 – STAR-3D-VT cohort results of 200 STAR-3D virtual patients compared to STAR-3D clinical data.

	STAR-3D-VT	**STAR-3D**	**P-value**
Number of patients	146	146	/
Number of episodes	200	200	/
Hours of control (h)	12201	12189	/
Total BG measurements	7279	6980	/
Workload (measurements per day)	14.3	13.7	/
Median [IQR] BG (mmol/L)	6.4 [5.8 7.4]	6.7 [6.0 7.7]	**<0.05***
% BG in 4.4-6.5 mmol/L	52	43	/
% BG in 4.4-7.0 mmol/L	67	60	/
% BG in 4.4-8.0 mmol/L	82	78	/
% BG in 8.0-10.0 mmol/L	10	13	/
% BG > 10.0 mmol/L	7	8	/
% BG < 4.4 mmol/L	0.9	0.9	/
% BG < 4.0 mmol/L	0.4	0.3	/
% BG < 2.2 mmol/L	0	0	/
Insulin rate (U/h)	5.0 [3.0 6.0]	4.0 [2.0 6.0]	**<0.05**
Dextrose rate (%GF)	100 [70 100]	98 [80 100]	**<0.05**

Data is given as median [IQR] where appropriate. Statistical testing using 95% CI bootstrapped BG medians difference. (*) indicates 95% CI bootstrapped BG medians difference is clinically equivalent.

Table 10.7 – STAR-3D-VT per-patient results of 200 STAR-3D virtual patients compared to STAR-3D clinical data.

	STAR-3D-VT	STAR-3D	P-value
Number of patients	146	146	/
Number of episodes	200	200	/
Episode length (days)	1.8 [0.8 3.5]	1.7 [0.8 3.5]	NS
BG measure/day per patient	14.4 [11.5 20.7]	13.9 [11.6 16.7]	NS
Initial BG (mmol/L)	10.5 [8.8 12.8]	10.5 [8.8 12.8]	NS
Median BG (mmol/L)	6.3 [5.9 6.9]	6.6 [6.2 7.2]	**<0.05[a]***
% BG in 4.4-6.5 mmol/L	56 [35 72]	45 [29 61]	**<0.05[a]**
% BG in 4.4-7.0 mmol/L	72 [51 84]	63 [41 78]	**<0.05[a]**
% BG in 4.4-8.0 mmol/L	85 [69 94]	81 [65 92]	**<0.05[a]**
% BG in 8.0-10.0 mmol/L	7 [3 16]	10 [5 18]	**<0.05[a]**
% BG > 10.0 mmol/L	2 [0 12]	4 [0 14]	NS
% BG < 4.4 mmol/L	0 [0 0.6]	0.0 [0.0 0.0]	NS
% BG < 4.0 mmol/L	0 [0 0]	0.0 [0.0 0.0]	NS
% BG < 2.2 mmol/L	0 [0 0]	0.0 [0.0 0.0]	NS
# patients < 2.2 mmol/L	0	0	1.00[b]
Insulin rate (U/h)	4.5 [3.0 6.0]	3.5 [2.5 5.0]	**<0.05[a]**
Dextrose rate for those fed (%GF)	95 [75 100]	97 [80 100]	NS

Data is given as median [IQR] where appropriate. Statistical testing using ([a]) 95% CI bootstrapped BG medians difference, or ([b])Fisher Exact test. (*) indicates 95% CI bootstrapped BG medians difference is clinically equivalent. NS = not statistically significant.

Per-episode, STAR-3D-VT provides significantly higher performance in all intermediate ranges ($p<0.05$), for similar safety, and clinically equivalent per-episode median BG levels (Table 10.7). While nutrition rates are similar, insulin rates are significantly higher ($p<0.05$) in simulations (4.5 [3.0 6.0] U/h) compared to clinically (3.5 [2.5 5.0] U/h). More importantly, workload was slightly higher, but similar, between STAR-3D-VT (14.4 [11.5 20.7] measures per day) and STAR-3D (13.9 [11.6 16.7] measures per day).

10.3.4. Compliance to Protocol

A protocol compliance analysis of STAR-3D was done to determine how often nurses chose the longest treatment interval suggested, and how often nurse changed the recommendations (Table 10.8). The longest treatment intervals of 1, 2, and 3 hours suggested by STAR-3D were 55%, 12%, and 33% of recommendations, respectively. Additionally, nursing staff selected the longest treatment interval option 96% of the time. Nurse compliance to protocol was extremely high, with only 7% insulin and 4% nutrition deviations from original recommendations. The CDFs of the difference in insulin and nutrition rates is shown in Figure 10.6, where ~80% differences in insulin rate are ±3U/h, and ~80% differences in nutrition rate are ± 5%GF.

Table 10.8 - Compliance analysis results for STAR-3D.

	STAR-3D
Median [IQR] max treatment interval available (mean)	1 [1 3] (1.8)
% max interval chosen	96
% insulin intervention unchanged	97
% nutrition intervention unchanged	96

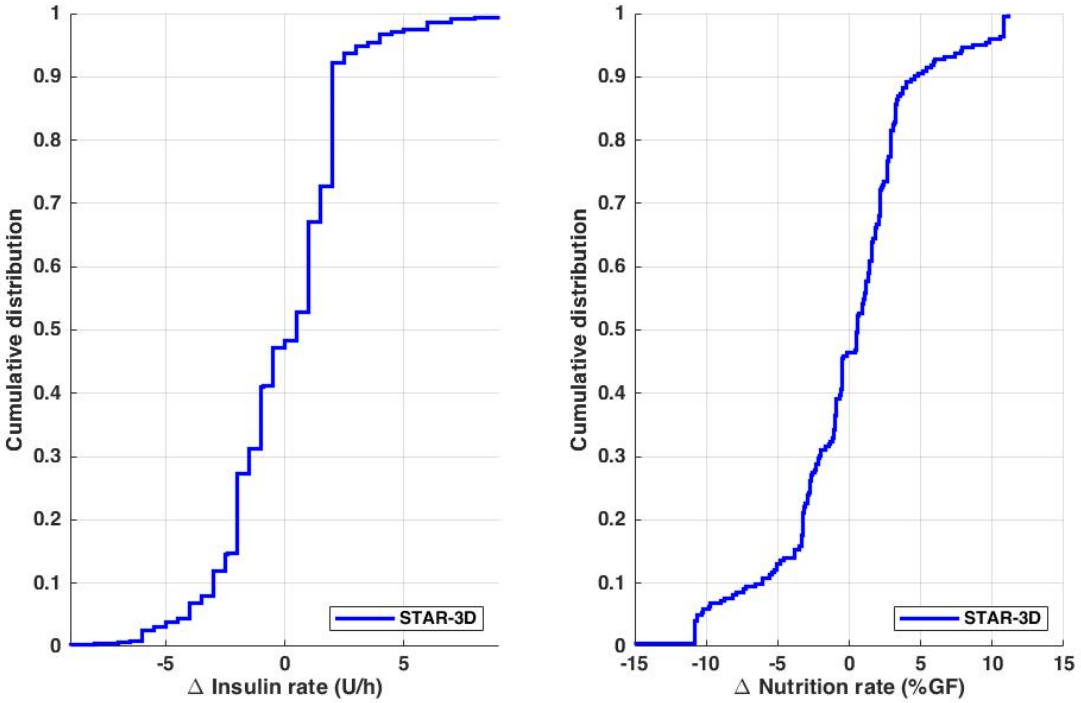

Figure 10.6 – Cumulative distribution functions of insulin (left) and nutrition (right) deviation from original recommendations. The difference in insulin rate is the sum of both bolus and background infusion deviations, explaining the potential difference of 0 U/h.

10.4. Discussion

Clinically, STAR-3D provided high, similar, GC safety and performance, especially in intermediate lower ranges compared to STAR-2D. High percentage time in those ranges, associated with improved outcomes [38, 76, 77]. The incidence of hyperglycaemia was higher in STAR-3D, explained by the higher initial starting BG levels (Table 10.3 and Figure 10.4). Hypoglycaemia was lower in STAR-3D, with no incidence of severe hypoglycaemia compare to STAR-2D, where severe hypoglycaemia is associated with increased morbidity and death [10, 32, 50].

In addition, the left panel in Figure 10.4 presents the ratio of BG levels over initial episode BG level, showing a clear shift to the left for STAR-3D. Given the higher initial BG seen in the right panel of the same Figure, it implies that in clinical practice, with the same starting BG levels, STAR-3D would likely have shown tighter or higher performance than STAR-2D.

To capture this gap and avoid the influence of higher starting BG, an analysis only calculating outcome results once the target band was achieved. Results showed STAR-3D, once the target band reached, provides safer and more effective control than STAR-2D, while achieving significantly higher nutrition rates. Thus, STAR-3D is able to achieve tighter control, with significantly higher feed rates, which can significantly improve patient outcome. The higher insulin and nutrition rates achieved compared to STAR-2D show the ability of STAR-3D to better personalise control according to patient-specific needs. Nutrition intake is safely optimised using more insulin. These outcomes are achieved with similar workload per-episode (Table 10.3 and Table 10.5).

Clinical staff compliance to protocol was extremely high. Not only was the longest treatment interval was chosen 96% of the time, but only 7% of insulin and 4% of nutrition recommendations were overridden. This result reflects high nursing compliance to protocol and thus good protocol design [152, 160, 174, 179]. It also suggests non-compliance had minimal impact on GC outcomes, and control was thus minimally influenced by nursing clinical judgment, where large negative (and positive) deviations in Figure 10.6 likely reflect STAR being turned off (restarted) to take patients to surgical or imaging procedures.

Comparing virtual trials results and clinical results of STAR-3D, there was improved performance in the STAR-3D-VT (Table 10.6 and Table 10.7). The resulting lower median BG achieved is associated with improved outcomes in ICU patients [38, 76, 77]. However, virtual trials represent ideal conditions with full protocol compliance. The reported significant lower insulin used clinically ($p<0.05$) could be a consequence of the less punctual timing of BG measurements. Equally, it may reflect impact of clinical judgment, and over-conservative approach fearing hypoglycaemia. More specifically, the 3D stochastic model provides tighter forward prediction ranges (more than 70% of the time, as presented in Chapter 9 [126, 127]), leading to more aggressive dosing of insulin compared to STAR-2D, as seen in Table 10.2 - Table 10.5.

As explained in Chapter 9, the net benefit from STAR-3D may be hard to capture clinically. If there is a clear increased insulin and nutrition rates, the BG levels achieved, the performance, and even safety are very similar here. However, the overall shift in BG seen and lower incidence of hypoglycaemia reflects the 3D stochastic model's ability to better predict intra-patient variability. The change from 1%

of patient of severe hypoglycaemia to 0% with STAR-3D over 200 episodes is especially important, even if it is not yet statistically significant, given the impact of severe hypoglycaemia on mortality [10, 11, 32, 33, 163]. Thus, the STAR's risk-based dosing approach is more personalised and tailored to patient metabolic evolution.

The 3D stochastic model better predicts extremes in SI variability, especially for more variable patients [126, 127]. Additionally, it also better accounts for measurement errors, as it is indirectly accounted for when building the model. STAR-3D is thus likely more robust to this error, and likely provide safer control due to improved prediction of variability. However, more patients should be included to analyse results on a broader cohort, and determine whether this new 3D stochastic model brings significant improvements on a larger cohort of patients.

This analysis presents intermediate results this ongoing STAR-3D clinical trial. GC statistics were thus easily available and comparable, but additional work should be undertaken to access more detailed demographic data, such as severity of injury, length of ICU stay, mortality, *etc*. This would determine whether both cohorts are comparable demographically, although they come from the same ICU here, a factor potentially impacting GC outcome.

10.5. Summary

In this analysis, clinical trial results of the STAR-3D protocol, using the new 3D stochastic model developed in Chapter 9, were compared to retrospective clinical data from the original version of STAR-2D. STAR-3D provided similar, high, quality of control while reducing the incidence of hypoglycaemia, despite using significantly higher insulin rate, and significantly increasing nutrition delivery compared to STAR-2D.

The use of a 3D stochastic model enables improved, personalised, prediction of metabolic variability, and thus improves GC outcome in the context of STAR. The results suggest the continuation of this trial, and further investigation to clearly determine and compare the clinical benefits associated with this new stochastic model.

Chapter 11: STAR-Liège Clinical Trial

This chapter presents clinical results of the ongoing STAR-Liège clinical trial at the University Hospital of Liège, Belgium. The STAR-Liège aims to assess safety and performance of STAR in a general ICU environment different to that in which it was developed, and compare results to local standards. This trial is used to further validate the generalisability of the STAR GC framework across different ICUs and practices.

Unique to this trial, it includes patients on a STAR Insulin-Only version (STAR-IO), leaving nutrition at clinician discretion. These results are compared to full STAR modulating both insulin and nutrition inputs, providing a first ever clinical quantification of the impact of modulating nutrition. This study thus also analyses the impact of modulating nutrition on GC outcomes, in the context of the proven STAR GC framework.

This chapter presents results partially published in [259], [154], and [260].

11.1. Introduction

As previously stated, GC to reduce BG levels to safer concentrations has shown improved outcomes, reducing organ failure, clinical burden, and costs [35-37, 171]. However, other studies failed to replicate these results [8, 48-50, 65, 66, 261], showing increased glycaemic variability and higher risk of hypoglycaemia, independently associated with severe complications and death [10, 11, 14, 32, 33, 55, 58]. To date, the optimal target band for GC is still being debated (Chapter 1) [60-62, 262].

Over the years, some critical factors have been identified for good protocol design, including safety, performance, compliance to protocol, and repeatability across units [86, 87, 91, 135, 152, 160]. However, few clinical settings, if any, evaluate and quantify these factors directly before implementation of a GC protocol in a clinical trial or as a standard of care. Safe, effective control must be achieved for all patients before potentially assessing its impact on clinical outcomes [64, 135]. Using a GC protocol design not providing all of these critical aspects can bias study results and conclusions [135].

STAR has shown positive results in two ICUs from New Zealand and Hungary, providing safe control for nearly all patients [87], where it is now the standard of care for GC. While all these independent before-after single centre trials suggest STAR is a successful, adaptable solution, the significance of each single study is underestimated versus the body of work as a whole. Results of STAR are thus not considered by reference studies updating GC guidelines in ICUs [70, 72].

The STAR-Liège clinical trial was implemented in the University Hospital of Liège, Belgium, to show STAR's ability to adapt to another local ICU's standards and practices. It is important to note there are many differences in the protocol design of this study with prior pilot trials of STAR in Belgium [149, 151]. First, the target band used here is lower (4.4-8.0mmol/L) compared to >6.9 mmol/L in previous studies, and nutrition was left at clinician discretion in one part of this study. Additionally, clinical staff for this clinical trial use STAR directly, without technical support. Results are thus more representative of the full usage of STAR by nurses, potential handling errors, and impact on ICU practices and nurse compliance.

STAR is not the only successful model-based protocol [43, 81, 108], but it is the only GC protocol also modulating nutrition for increased quality of control, while optimising CHO intake [263], a significant point of difference from these protocols, and most, if not all, published clinical protocols. This analysis thus also quantifies, for the first time, the impact of modulating nutrition in addition to insulin on GC outcomes, in the context of the proven STAR GC framework (Chapter 3), by analysing two separate arms of this clinical trial.

This chapter thus presents and compares safety, performance, and compliance clinical trial results of STAR and a modified, insulin-only version of STAR (STAR-IO), leaving nutrition at clinician discretion. The University Hospital of Liège Ethics Committee approved this trial (#**B707201733994**) and the use of the data presented.

11.2. Methods

11.2.1. STAR-Liège Clinical Trial Design

The STAR-Liège clinical trial was designed to include two arms of 20 patients each. In the first arm (STAR-IO), only insulin is modulated, and nutrition, while known, is left at clinical discretion. In the second, the full version of STAR is used, modulating both insulin and nutrition inputs (Table 11.1). Informed consent is collected from all patients, which can be freely withdrawn at any time during the trial.

STAR is fully computerised and deployed on Android™ operating system tablets, and easily adjusts to local ICU practices. Nurses enter BG, insulin, and nutrition data directly in the tablet. STAR then operates using the patient data to compute a new personalised treatment, using the unique risk-based dosing approach explained in Chapter 3 [95, 96, 123].

ICU patients are included in the study if they have an intra-venous (IV) catheter, are enterally fed, have a survival prognosis of minimum 72h, and are in the need of insulin therapy (two consecutive BG measurements > 8.0 mmol/L). The STAR target band is the normoglycaemic range of 4.4-8.0 mmol/L. Insulin is continuously administered through intra-venous catheter, with a maximum insulin rate of 9U/h. Increments of a maximum +2U/h are allowed in the infusion rate between successive interventions.

Nutrition, in the full STAR version, can be temporarily decreased down to a minimum of 30% of the original clinically set 100% GF. Typically, nutrition is reduced if insulin alone is not sufficient to decrease persistent elevated BG levels. Nutrition can only be reduced by a maximum 30% between consecutive measurements. In STAR-IO, nutrition is known to STAR, but the rate is left at clinician discretion. In the case of hypoglycaemia (BG<3.0 mmol/L), a dextrose bolus (20ml of 30% glucose) is administered intravenously while insulin is stopped. A new BG measurement will be needed within one hour.

STAR stopping criteria are BG levels stable (in target band) for 6 hours at low insulin rates (≤2U/h), or after 72 hours of control. If a patient still requires GC at 72h, they revert to standard practice GC. BG measurements are taken 1-3 hourly based on STAR recommendations and nursing choice [95, 96, 123]. All assays are measured using a blood gas analyser (GEM Premier 5000™, Instrumentation Laboratory, USA).

11.2.2. Standard Protocol

Clinical trial results are compared to retrospective data from the local standard GC protocol (SP, Table 11.1). The SP is a table-based protocol targeting 5.6-8.3 mmol/L. BG measurements are typically taken 4-hourly when 5.6 mmol/L < BG < 10.0 mmol/L, and 1-hourly otherwise. There is no specified maximum insulin infusion rate. Starting criteria requires one BG > 10.0 mmol/L, which is much higher compared to STAR. BG measurements are made using glucometers or a blood gas analyser. In the case of nutrition stoppage, insulin is automatically stopped. Otherwise, insulin administration is stopped only when BG is below 3.3 mmol/L, and a 20ml of 30% glucose bolus is administered for severe hypoglycaemia (BG < 2.2 mmol/L). The full SP design is described in [264, 265].

Table 11.1 - Summary of the STAR, STAR-IO, and SP protocol designs compared in this analysis.

	STAR	STAR-IO	SP
Type	Model-based	Model-based	Table-based
Target band	4.4-8.0 mmol/L	4.4-8.0 mmol/L	5.6-8.3 mmol/L
Measurement intervals	1-3 hourly	1-3 hourly	1 or 4 hourly
Insulin strategy	Max 9U/h	Max 9U/h	Not specified but ~15U/h clinically
Nutrition strategy	Modulation between 30-100% GF	Clinical discretion	Clinical discretion

11.2.3. Protocol Comparison and Analysis

To date, 15 patients were included under STAR-IO and 14 patients under the full STAR protocol. Results from 20 retrospective patients under SP published in [128, 264, 265] are used for comparison. Basic demographics are presented in Table 11.2.

Table 11.2- Demographic characteristics of patients included in STAR and STAR-IO, and retrospective SP patients.

	STAR	STAR-IO	SP
# patients	14	15	20
# control hours	690	803	5006
% Male	64	73	45
Age (years)	72 [61, 75]	72 [66, 78]	66 [52 72]

Data given as median [IQR] as appropriate.

Safety, performance, nutrition, workload, and compliance are compared. Safety is assessed by the percentage BG in mild and severe hypoglycaemia (BG ≤ 4.0 mmol/L and BG ≤ 2.2 mmol/L, respectively), and in severe hyperglycaemia (BG > 10.0 mmol/L). Performance is evaluated by the percentage BG in target band (4.4-8.0 mmol/L or 5.6-8.3 mmol/L), and per-patient median BG achieved. Workload considers the number of measurements per day, both for the cohort and per-patient. Nutrition comparisons are made using the per-patient dextrose rates achieved in g/h and in percentage of the original GF. BG measurements are resampled hourly to allow fair comparison of the data, using linear interpolation [122, 266].

For the compliance analysis, similarly to Chapter 10, the percentage of insulin and nutrition interventions unchanged from the original protocol recommendations are assessed. Only changes in insulin or nutrition rates occurring within 15 minutes after treatment selection are considered as a deviation from the original recommendation, where later changes are assumed due to clinical needs. The size of these deviations from protocol prescribed administration rates are also reported. Deviations from nutrition administration are analysed only for STAR, as nutrition for STAR-IO (and SP) was left to clinical discretion.

11.3. Results

11.3.1. STAR vs. STAR-IO

Whole cohort and per-patient clinical results for STAR and STAR-IO are presented in Table 11.3 and Table 11.4. Resampled BG, insulin, and nutrition CDFs are shown in Figure 11.1. Similar BG levels (6.7 [6.1, 7.5] mmol/L for STAR and 6.6 [5.9 7.6] for STAR-IO) are achieved by both protocols, but with overall tighter BG levels for STAR (Figure 11.1). Time in target band (4.4-8.0 mmol/L) is higher for STAR (83%) than STAR-IO (78%).

Safety was high and similar across both protocols, although STAR was safer, with only 0.4% moderate hypoglycaemia (BG < 4.0 mmol/L) compared to STAR-IO (0.7%). No patients in any arms experienced severe hypoglycaemia. On the other extreme, severe hyperglycaemia (BG > 10.0 mmol/L) is significantly lower for STAR (4%), compared to STAR-IO (9%), but mild hyperglycaemia (BG in 8.0-10.0 mmol/L) are similar (13% for STAR and 12% for STAR-IO).

Slightly higher insulin rates were administered for STAR (3.5 [2.4, 4.9] U/h) compared to STAR-IO (3.1 [1.5, 5.7] U/h), while dextrose rate was higher for STAR-IO (7.8 [5.1, 8.5] g/h) compared to STAR (6.9 [5.4, 8.3] g/h). However, this difference becomes smaller when considering nutrition based on %GF, calculated based on g/hr of nutrition per body weight and energy demand (Chapter 7), where STAR achieved more consistent cohort feeding rates compared to STAR-IO (98 [67, 109] %GF for STAR vs. 93 [53, 103] %GF for STAR-IO).

Per-patient, median GC episode length is lower for STAR (2.0 [0.8 3.0] days) than STAR-IO (2.3 [1.2 2.9] days), but workload is significantly higher in STAR-IO (15.9 [12.6 18.3] measures per day) than STAR (13.8 [10.9 14.5]) likely due to the greater incidence of BG > 10.0 mmol/L. Median BG levels achieved are similar (6.5 [6.3 6.9] mmol/L for STAR and 6.5 [6.2 6.8] for STAR-IO], but tighter in STAR (Figure 11.2). High, consistent, per-patient performance is similarly achieved in both cohorts for all BG intermediate ranges (Table 11.4). However, there is significantly higher incidence of hyperglycaemia in STAR-IO (7 [0 13] median %BG > 10.0 mmol/L) compared to STAR (0 [0 7] %BG >10.0 mmol/L), but also higher initial BG levels for STAR-IO (10.1 [9.3 12.8] mmol/L) compared to STAR (9.1 [8.9 11.1] mmol/L). Safety is similarly excellent in both protocols, with very near to zero hypoglycaemia for all

patients. Finally, these GC outcomes are achieved with slightly higher per-patient median insulin rates for STAR (3.5 [2.5 4.5] U/h) compared to STAR-IO (3.0 [2.0 4.5] U/h), and consistently higher, and closer to the 100% patient-specific energy expenditure, nutrition rates for STAR (95 [83 105] %GF) than STAR-IO (93 [54 122] %GF).

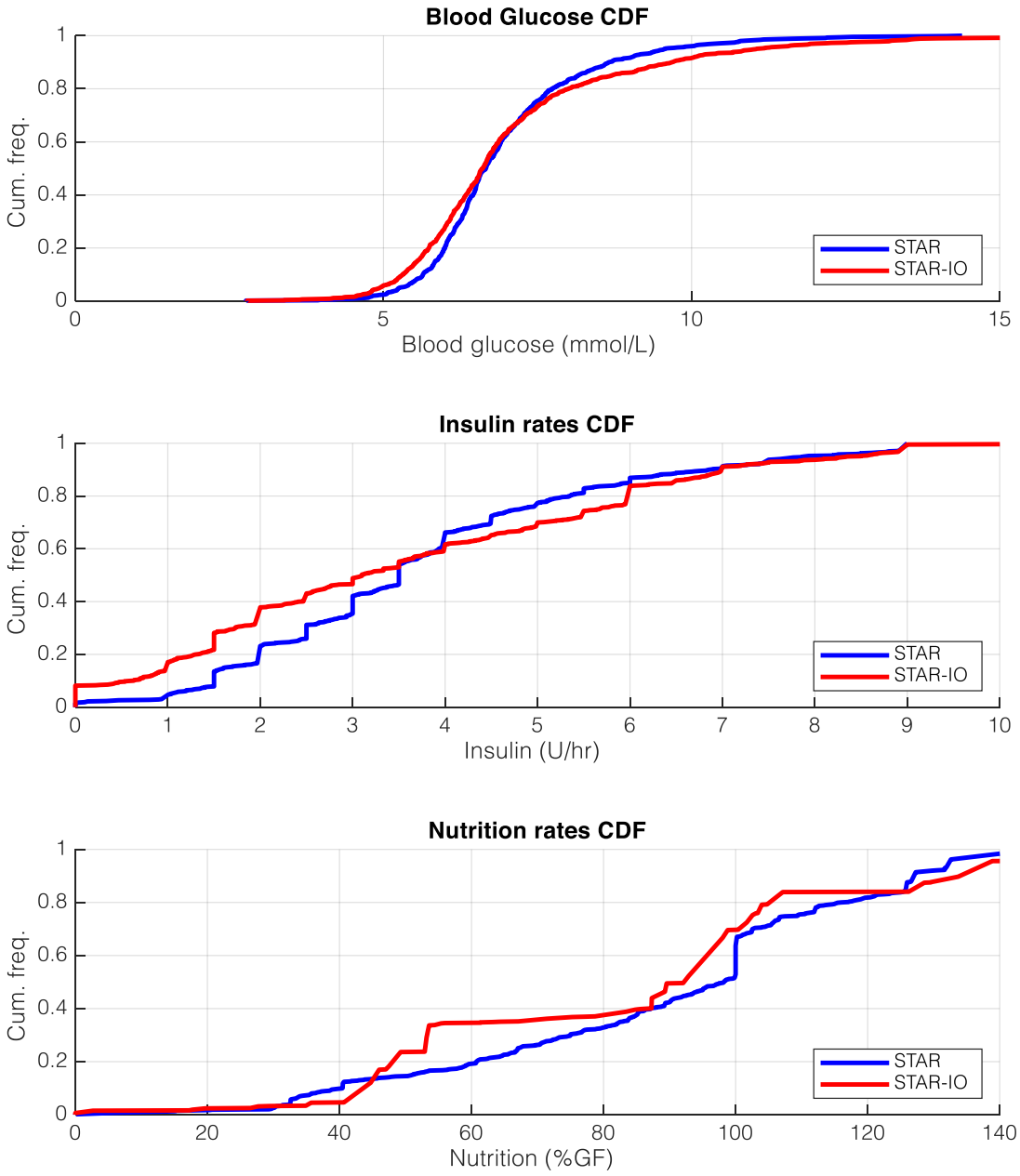

Figure 11.1 – BG, insulin, and nutrition cumulative distribution functions for STAR and STAR-IO.

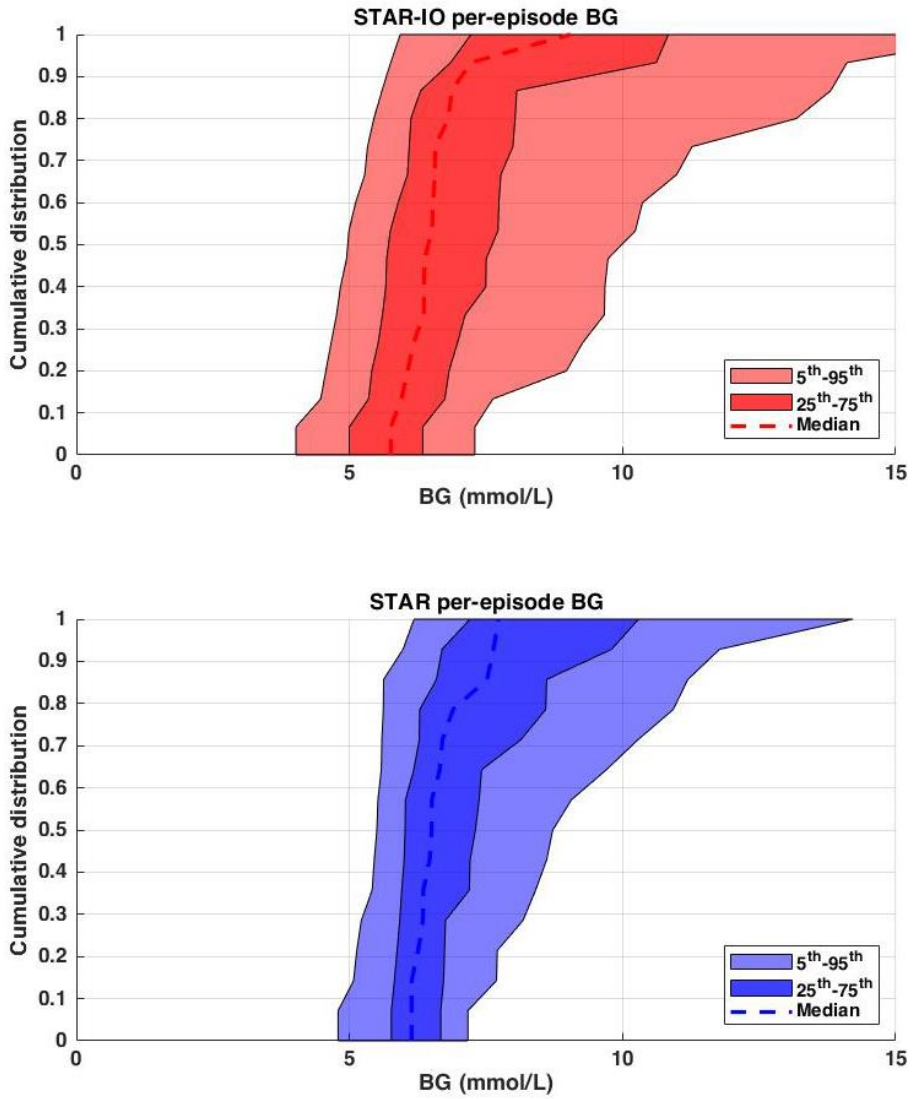

Figure 11.2 - Per-patient median BG as well as the 25th-75th and 5th-95th percentile ranges for STAR-2D (top) and STAR-3D (bottom).

Table 11.3 – Clinical cohort results for STAR, STAR-IO, and SP.

	STAR	STAR-IO	SP
# Patients	14	15	20
Total hours	690	803	5006
Total BG measurements	391	518	1391
Workload (BG measures/day)	13.6	15.5	7
Median BG (mmol/L)	6.7 [6.1, 7.5]	6.6 [5.9, 7.6]	7.7 [6.5, 8.9]
% BG in 4.4-6.5 mmol/L	42	43	/
% BG in 4.4-8.0 mmol/L	83	78	55
% BG in 5.6-8.3 mmol/L	78	67	54
% BG in 8.0-10.0 mmol/L	13	12	31
% BG > 10.0 mmol/L	4	9	12
% BG < 4.4 mmol/L	0.7	1.5	1.3
% BG < 4.0 mmol/L	0.4	0.7	0.5
% BG < 2.2 mmol/L	0	0	0
# Patients < 2.2 mmol/L	0	0	0
Insulin rate (U/h)	3.5 [2.4 4.9]	3.1 [1.5 5.7]	2.5 [2.0 3.0]
Dextrose rate for hours fed (%GF)	98 [67 109]	93 [53 103]	/
Dextrose rate (g/h)	6.9 [5.4 8.3]	7.8 [5.1 8.5]	9.7 [8.8 11.7]

Data given as median [IQR] as appropriate.

Table 11.4 – Per-patient clinical results for STAR, STAR-IO and SP.

	STAR	STAR-IO	SP
Number of patients	14	15	20
Episode length (days)	2.0 [0.8 3.0]	2.3 [1.2 2.9]	8.2 [4.9 12.4]
Workload (BG measures/day)	13.8 [10.9 14.5]	15.9 [12.6 18.3]	/
Initial BG (mmol/L)	9.1 [8.9 11.1]	10.1 [9.3 12.8]	8.5 [7.3 9.9]
Median BG (mmol/L)	6.5 [6.3 6.9]	6.5 [6.2 6.8]	7.8 [7.3 8.2]
% BG in 4.4-6.5 mmol/L	48 [33 58]	45 [38 55]	/
% BG in 4.4-8.0 mmol/L	88 [64 89]	75 [71 88]	/
% BG in 5.6-8.3 mmol/L	81 [63 86]	65 [57 72]	/
% BG in 8.0-10.0 mmol/L	11 [5 18]	11 [6 18]	/
% BG > 10.0 mmol/L	0 [0 7]	7 [0 13]	/
% BG < 4.4 mmol/L	0.0 [0.0 1.1]	0.0 [0.0 3.2]	/
% BG < 4.0 mmol/L	0.0 [0.0 0.0]	0.0 [0.0 1.6]	/
% BG < 2.2 mmol/L	0.0 [0.0 0.0]	0.0 [0.0 0.0]	/
# patients < 2.2 mmol/L	0	0	0
Insulin rate (U/h)	3.5 [2.5 4.5]	3.0 [2.0 4.5]	2.7 [2.0 3.0]
Dextrose rate for those fed (%GF)	95 [83 105]	93 [54 122]	/
Dextrose rate (g/h)	6.7 [6.3 8.2]	7.8 [6.2 8.4]	9.8 [8.6 11.5]

Data given as median [IQR] as appropriate.

11.3.2. STAR and STAR-IO vs. SP

Available data from retrospective analysis of patients under the SP [264, 265] are also presented in Table 11.3 and Table 11.4 for comparison with STAR. Clearly, STAR and STAR-IO provide significantly higher performance than SP, much higher %BG in both the STAR target band (>78%BG in 4.4-8.0 mmol/L for STAR and STAR-IO vs. 55% for SP) and the SP target band (>67% vs. 54% BG in 5.6-8.3 mmol/L, respectively). Accordingly, the median BG achieved in SP is higher (7.7 [6.5, 8.9] mmol/L).

Moderate hyperglycaemia (BG in 8.0-10.0 mmol/L) is significantly higher for SP (31%) compared to STAR and STAR-IO (<13%), as well as severe hyperglycaemia (12% BG >10.0 mmol/L for SP vs <9% for STAR and STAR-IO). Importantly, these lower performing results are achieved despite lower per-patient initial BG (8.5 [7.3 9.9] mmol/L).

More importantly, safety is similar (~0.5% <4.4 mmol/L) for all protocols, despite STAR and STAR-IO targeting lower BG ranges. This level of safety is achieved with lower insulin rates (2.5 [2.0 3.0] U/hr) and higher nutrition rates (9.7 [8.8 11.7] g/hr) for SP compared to STAR (3.5 [2.4 4.9] U/h and 6.9 [5.4 8.3] g/hr) and STAR-IO (3.1 [1.5 5.7] U/h and 7.8 [5.1 8.5] g/h). Finally, workload is much lower in SP (7 measurements per day) compare to STAR and STAR-IO (~14-15 measurements per day), as expected from protocol design (4-hourly for SP vs. 1-3 hourly for STAR).

11.3.3. Compliance Analysis

Results from the compliance analysis are presented in Table 11.5. Compliance to protocol was high in both STAR and STAR-IO. Clinical staff overrode insulin administration rates in only 13% of all treatments (49/373 overrides) for STAR, and 8% (42/503 overrides) for STAR-IO. As shown in Figure 11.3, from those deviations, 60% were reduced by nurses in STAR compared to 50% in STAR-IO. In STAR, nutrition rate recommendations were only changed 13% (48/373 overrides) of the time. Under the SP, there was a higher 21% deviation from original protocol. As reflected in the lower workload for STAR (Table 11.3), STAR more often suggested longer treatment intervals compared to STAR-IO (1 [1 3] h (mean 2.0h) vs. 1 [1 2] h (mean 1.6h), respectively). In both arms, the maximum treatment was selected by clinical staff ~90% of the time.

In total, there are 14 (48%) GC episodes with non-compliance, of which 6 (43%) are patients included on STAR and 8 (57%) included on STAR-IO. GC episodes with non-compliance mainly occurred during the inclusion of the first third of patients, where 70% (7/10) had non-compliance, compared to 37% (7/19) for the last two thirds of patients. Thus, GC non-compliance is dropping, likely reflective of clinical staff protocol uptake and education.

Table 11.5 – Compliance analysis results for STAR and STAR-IO, and the SP

	STAR	STAR-IO	SP
Max treatment interval available (mean)	1 [1 3] (2.0)	1 [1 2] (1.6)	(3.65)
% max interval chosen	90	91	/
% insulin intervention unchanged	87	92	79
% nutrition intervention unchanged	87	/	/

Data given as median [IQR] as appropriate.

11.4. Discussion

Clinical results from this ongoing trial are encouraging, and suggest key observations. Overall, STAR (and STAR-IO) achieved safe and effective control for all patients, despite targeting a lower target band than SP or those usually recommended in ICU guidelines. This result suggests intensive GC to lower target bands is possible without increasing hypoglycaemic risk. Furthermore, it reinforces the idea GC has been wrongly blamed for hypoglycaemia [160], while protocol design and GC approach are the primary concerns to safely achieve high quality GC outcomes. This goal is essential before assessing potential clinical outcome, and failing to do so would suggest poor protocol design [135].

The difference in %BG in the 4.4-8.0 mmol/L range using STAR and STAR-IO (~80%) compared to SP (55%) is significant, and these ranges have been associated with improved outcomes in numerous studies [38, 76, 77, 79]. While this result could be explained from the modestly different target band, SP only managed to have 54% BG within its target (5.6-8.3 mmol/L), where STAR (78%) and STAR-IO (67%) performed better in this range also.

Figure 11.2 clearly shows per-patient tighter BG levels achieved in STAR compared to STAR-IO, where median BG CDFs are similar, but BG CDFs of the 25th-75th and 5th-95th are wider. The higher severe hyperglycaemia in STAR-IO compared to STAR is reflected in the 95th percentile per-patient BG CDF shift to the right, showing >50% of STAR-IO patients had 5% BG >10.0mmol/L. However, this difference can be explained, at least in part, by the higher initial BG for STAR-IO, or equally, the significantly higher nutrition rates provided by the clinical staff in STAR-IO (25% of patients receiving >120%GF in STAR-IO, Table 11.4).

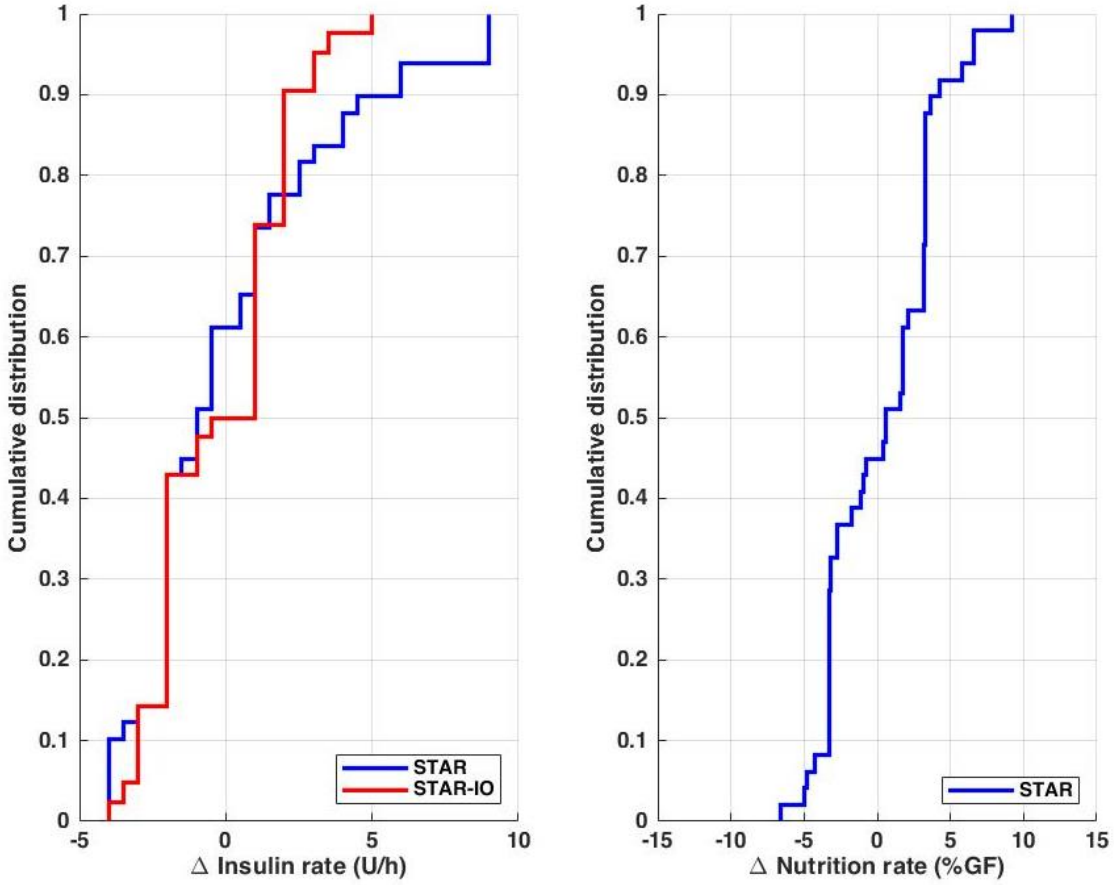

Figure 11.3 – Cumulative distribution function of insulin (left) and nutrition (right) deviation from protocol recommendations. Differences are computed so that a negative Δ corresponds to a reduction from the original recommendation, while a positive Δ corresponds to an increase from the original recommendation.

For the SP, percentage BG in mild hyperglycaemia is high compared to STAR and STAR-IO (31% vs. ~12%). The higher target could somewhat explain this result, or the higher BG level starting criteria (10.0 mmol/L compared to 8.0 mmol/L with STAR), but the initial BG levels in SP were unexpectedly much lower than STAR and STAR-IO, amplifying the gap in performance. However, it is likely a consequence, in part, of clinical judgement considering BG in 8.0-10.0 mmol/L under the SP protocol as acceptable. The lower compliance to protocol for SP (79%) could also explain some of this result (Table 11.5).

A previous analysis showed 68% of the 21% total insulin interventions changed from original SP recommendations were made when BG was above target band [265]. For those 68%, nurses (unexpectedly) decreased insulin in 62% of the deviations. While in band, 18% of the 21% total intervention changes were made, from which 78% were a decrease in insulin rate. These changes clearly suggest potential shift to higher BG levels due to clinical judgement.

Protocol compliance for STAR (87%) and STAR-IO (92%) was high in this clinical trial. Off these 13% (49) insulin deviations in STAR, 80% (39) were made when in target band, 20% (10) when above target, and none when below target. In STAR-IO, 71% (30), 24% (10), and 5% (2), respectively, from the total 8% (42) of insulin intervention changed. There is no clear pattern to the way insulin rates were adjusted. However, observations of large insulin deviations (larger than ± 2.5U/h) suggest bedside clinical staff were not able to balance insulin against nutrition rates, which is automatically considered by STAR. Most of the large modulations of insulin to higher rates occurred when BG was increasing. In this context, STAR typically recommended a reduction in nutrition rates, and, thus, lowered insulin to balance the risks of hypoglycaemia. In contrast, clinical staff often kept insulin constant and thus very high relative to nutrition. In other cases, a decrease in BG was often accompanied by an increase in recommended nutrition, and maintenance of a constant or increased insulin dose to match. Overall, it seems that clinical staff struggle to manage, assimilate, and mitigate current BG level, risk of hypoglycaemia, any glycaemic variability, and the relative changes in insulin and nutrition into appropriate bedside action. Most often, it seems clinical staff had difficulty to take into account CHO intake, and balance that against insulin administration. These deviations highlight the benefit of model-based systems, which can account for multiple variables when calculating treatments. In the worst case, this saturation of variables in clinical decision making can result in safety issues, such as in the case of overriding and increasing the insulin dose despite lowering nutrition.

STAR's lower incidence of mild and severe hyperglycaemia compared to STAR-IO is a consequence of the combination of nutrition modulation, and higher initial BG. This result is also reflected in the overall nutrition rates achieved, and insulin requirements (Figure 11.1). While overall nutrition is lower in STAR, the gain in performance is significant. In fact, nutrition below GF for STAR is minimal in Table 11.3 because nutrition is mainly temporarily decreased for very resistant patients, where BG remains high while receiving the maximum insulin rate. These patients' BG levels can only be lowered if glucose intake is lowered [263]. Interestingly, there was 25-30% of control hours where nutrition rates were higher than STAR 100%GF recommendations. This augmented rate is due to the background parenteral feed rates often administered to patients in this ICU. A recent study analysed the nutrition delivery of STAR compared to other ICUs in the world, and showed STAR performs equal to the best ICUs in the world [134]. Therefore, despite modulating nutrition, STAR does not underfeed patients, and still manages to improve GC outcomes.

A case study is shown in Figure 11.4, where glucose-insulin interventions and outcomes are shown for a patient on STAR and another on STAR-IO. In Patient A, on STAR, nutrition is first increased to 100% GF then reduced to ~30% after 6 hours GC. Underlying SI allows STAR to reduce nutrition rates, where excess nutrition will not be tolerated, as would be indicated by persistent hyperglycaemia. Hence, this temporary lower dextrose rate is safer for the patient. By hour 31, the nutrition rate was progressively increased back to 100% GF as the patient metabolism was better able to handle higher dextrose intake. In contrast, nutrition rate is kept constant for Patient B, as per clinical guidelines, resulting in consistent high BG levels, above, or close to the upper target band limit, with higher associated insulin rates. This higher fixed nutrition and associated high insulin rate may often lead to hypoglycaemia and higher glycaemic variability if patient SI changes due to underlying condition, or other clinical interactions.

These results are achieved with much higher workload with STAR (13.6 measures per day) and STAR-IO (15.5 measures per day) compared to SP (7 measures per day). This difference reflects the big trade-off of measurement frequency to achieve high control quality, where too low a frequency can increase hypoglycaemic risk and lower time in band, while too high a frequency leads to excessive clinical burden. The resulting median [IQR] BG is much higher than STAR, with 25% of BG measurements higher than 160 mg/dL.

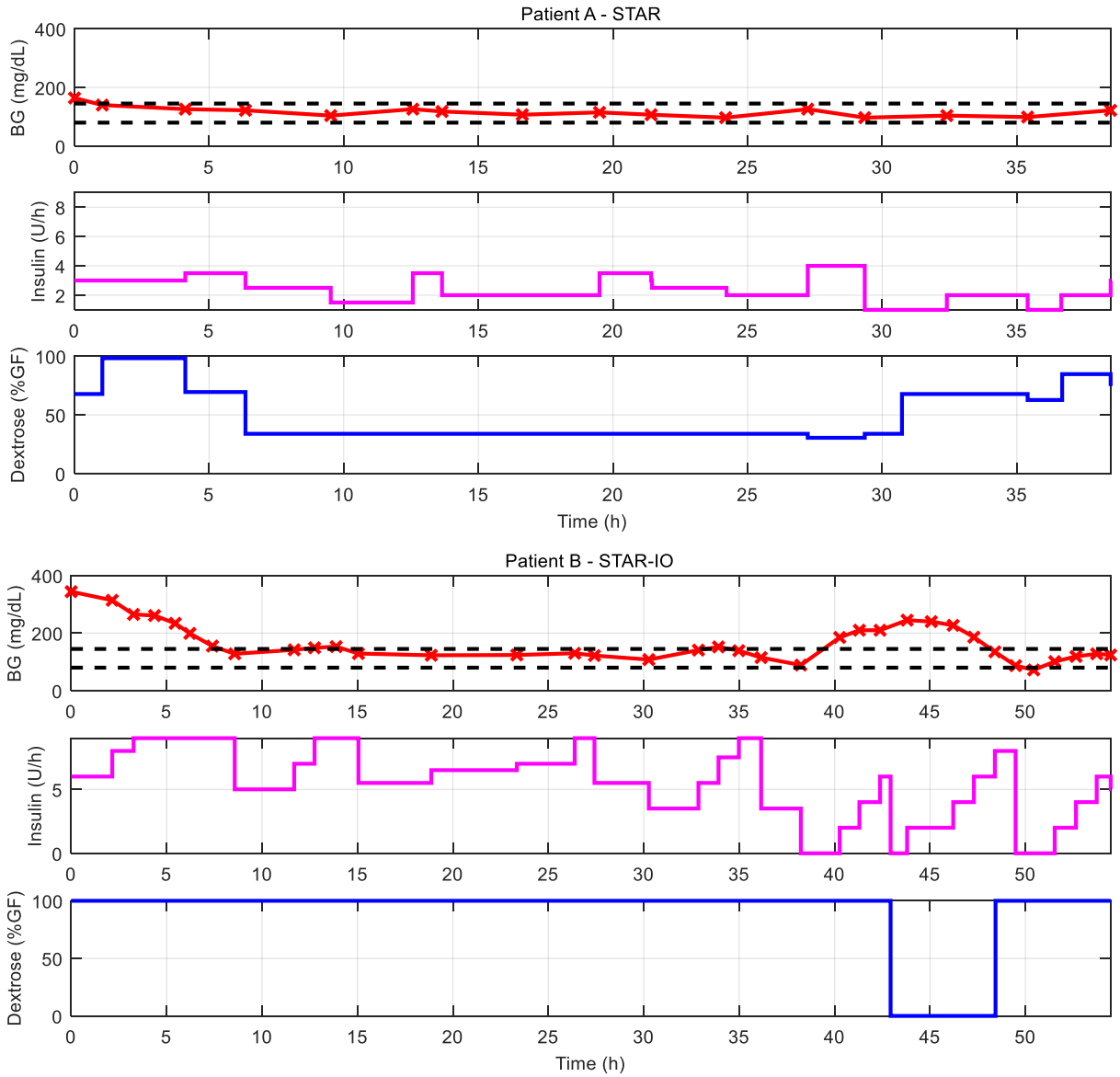

Figure 11.4 – BG levels, insulin rate, and nutrition rate for Patient A (top) on STAR, and Patient B (bottom) on STAR-IO. Dashed line shows the 4.4-8.0 mmol/L target band.

While being the standard of care in this ICU, more than 20% of SP interventions are changed by clinical staff, leading to a large and important influence from clinical judgment on GC outcomes, which in turn questions the results and protocol design. Additionally, in a previous study, virtual trials of SP on virtual patients created using this cohort data suggested, despite similar GC outcomes, a likely low compliance to protocol [259]. Simulations show SP needed an average of 11 measurements per day when exactly following the protocol, much higher than the 7 observed here, and, more importantly, much closer to the 13.6 measurements per day required by STAR. This outcome thus suggests, in addition of higher deviation from protocol recommendations, the measurement frequency was also an issue impacting compliance to protocol, and, thus, GC outcome.

It is also important to further note the important difference in workload, at a per-patient level, between STAR and STAR-IO. STAR workload (13.8 [10.9 14.5] measures per day), was much lower compared to STAR-IO (15.9 [12.6 18.3] measures per day), with an average extra 2-3 measures per day required. Per-patient workload CDFs are presented in Figure 11.5 for both protocols, clearly showing ~50% of patients required ≥~16 measures per day, compared to ~15% for STAR. Although initial BG was higher in STAR-IO, this important difference is most probably a consequence of excessive nutrition administered to highly resistant patients, resulting in excessive BG levels, and thus more interventions required to attempt reaching the target.

Compared to 2 published pilot trials of STAR (SL1 and SL2) in the same Hospital [77, 267], each including 9 patients, STAR and STAR-IO achieved significantly safer, more effective control. In SL1 and SL2, nutrition was left at clinician discretion. Additionally, workload was extremely high in those trials (>17 BG measurements per day), a clear clinical workload failure. Finally, nurse compliance was ~75% in SL2, much lower than for STAR and STAR-IO (Table 11.5). Hence, the approach, and thus compliance and control quality, have all improved in this study.

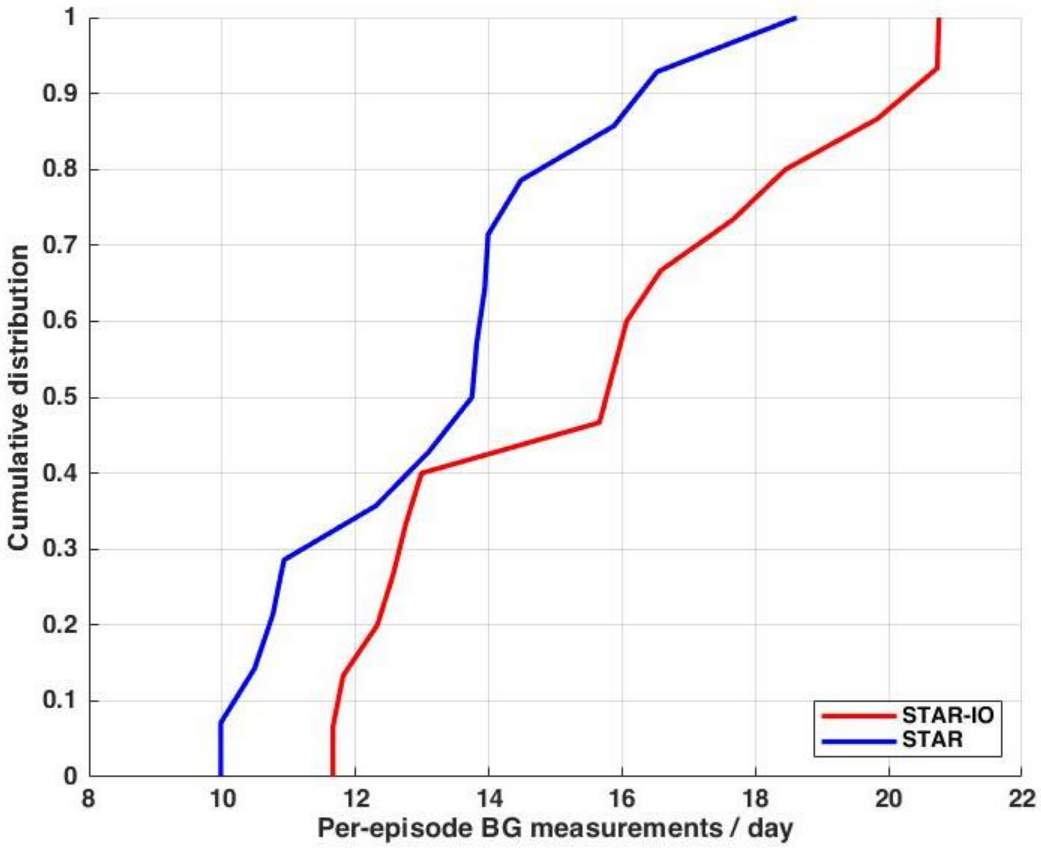

Figure 11.5 – Per-patient workload CDF for STAR-IO (red) and STAR (blue).

The main difference between STAR and STAR-IO is the ability to also modulate nutrition based on the identified patient-specific ability to take up BG (SI), and resulting risk-based assessment associated with a given treatment. Hence, STAR is not only able to minimise hypoglycaemia, but it can also, in addition, minimise hyperglycaemia by temporarily turning down caloric intake. The results of this trial comparting both protocols clearly show STAR's ability to significantly improve GC performance, while ensuring similar (or better) safety and, also, providing improved personalised nutrition compared to STAR-IO. In addition, in the context of this proven GC framework, modulating nutrition also significantly reduced workload, as STAR can better manage highly resistant patients by significantly reducing caloric intake until these patients can better assimilate energy expenditure.

This study compares clinical data to retrospective patients and has some limitations. The number of patients in each arm are not identical, and the results are not based on the exact same underlying cohorts e.g. Age. GC episode length is also extremely higher in SP (8.2 [4.9 12.4] days) compared to STAR (2.0 [0.8 3.0] days) and STAR-IO (2.3 [1.2 2.9] days). STAR has lower shifted distribution by 15-50% of time on GC compared to STAR-IO, where greater patient number may better delineate this trend. However, these patients are from the same general medical ICU, and the cohorts are believed to be representative of the overall population. Importantly, more patients should be included to generalise and further confirm these results, although there already is a clear, significant, improvement in GC outcomes compared to the SP.

This model-based GC protocol identifies and directly uses inter- and intra- patient variability to improve safety and efficacy of GC, avoiding reliance on clinical judgment [91]. Altogether, the improved safety and performance, associated with lower mortality, lower morbidity, and lower ICU LOS [11, 38, 45, 54, 55, 121, 161], might be worth the slightly increased workload. More importantly, STAR can adapt to local ICU standards and practice, and its insulin-only version (STAR-IO) still managed to provide safe and effective control for nearly all patients. The clinical trial results presented here thus further validate STAR's ability to provide high quality of control, and generalise internationally.

11.5. Summary

STAR was able to achieve safe and effective GC, while targeting lower intermediate glycaemic ranges compared to the local standard protocol, associated with improved outcomes. The full STAR version is also able to better tailor nutrition needs for the patient, by temporarily reducing caloric intake for persistent hyperglycaemia. This approach significantly improves GC safety, and efficacy compared to the insulin-only version (STAR-IO), for lower workload. These intermediate results of the STAR-Liège clinical trial are encouraging, and suggest the continuation of this trial.

The results presented in this chapter once again suggest reconsidering GC guidelines. GC needs to be safe and effective for all patients, regardless of patient condition. Computerised model-based methods using key physiological parameters to identify patient-specific needs are proving positive results in GC targeting lower glycaemic ranges.

While proving the high benefits associated with STAR, the associated increased workload compared to the SP can potentially be seen as an excessive clinical burden in ICUs with lower nursing staff, preventing uptake and use. This outcome suggests assessing the impact of longer treatment intervals on GC outcome, to potentially reduce workload is necessary. Understanding and quantifying explicitly this trade-off in the context of the STAR GC framework would be beneficial to its uptake and use, and allow an explicit, quantified clinical choice in how STAR is deployed and used.

Chapter 12: Longer Treatment Intervals, the Risk and Reward Trade-off

Compliance to protocol is essential for GC. While more frequent measurements enable safer GC, it often leads to excessive clinical burden. In turn, clinical staff may not fully follow protocol recommendations if burden is too high [179], impacting GC outcome. Non-compliance can potentially significantly bias results and conclusions, as shown in Chapter 5 [160].

STAR offers 1-3 hourly treatment intervals based on patient-specific risk assessment of hypoglycaemia, averaging ~12 measurements per day in Christchurch, New Zealand, and ~13 measurements per day in Liège, Belgium (Chapter 11). While some protocols use even lower measurement intervals, most protocols typically use 4-hourly treatment intervals [8, 85, 268], significantly reducing workload. However, these longer intervals are often associated with significant reductions in both the safety and efficacy of GC.

This chapter aims to assess the impact of increasing treatment intervals in the context of STAR on GC outcomes. Stochastic models predicting SI variability up to 6 hours are developed. Virtual trials are used to assess safety and performance using these longer measurement intervals within the STAR framework, thus quantifying the risk of reduced safety and performance versus the reward of decreased workload. This analysis is the first of its kind in GC.

This chapter presents results published in [269].

12.1. Introduction

To date, STAR uses 1-3 hourly forward prediction intervals to assess potential risk of hypo- and hyper-glycaemia for given 1-3 hourly treatments, averaging 11-12 BG measurements per day [87, 95, 96, 123]. While some ICUs can manage this workload, this value can be seen as excessive clinical burden for others, often due to lower nurse per patient ratios or greater clinical complexity of the patients. In the STAR-Liège clinical results (Chapter 11), the excessive workload compared to the standard protocol was clearly pointed out as a barrier to adopt this protocol. Equally, many clinical studies used longer intervals, but could not deliver safe, consistently effective GC [46, 47, 50, 89, 255].

This chapter extends from 1-3 hourly to 1-6 hourly measurement and intervention intervals in the STAR GC framework, and analyses the impact on GC safety and efficacy using clinically validated virtual patient modelling approach [91, 128]. If accomplished with minimally reduced safety and performance, this change has the potential to significantly reduce nurse workload, which is a major issue in GC [174, 270, 271]. It would also extend STAR's capability while increasing its acceptability for clinical use in more ICUs. More specifically, this study aims to assess the risk and reward trade-off associated with lower BG measurement frequency.

12.2. Methods

12.2.1. STAR 1-6 Hourly Extension

STAR currently uses 1-3 hourly measurements to provide GC [95, 96, 123]. This interval was originally chosen based on Christchurch (New Zealand) ICU standards and conservative decisions to ensure high safety and efficacy [81, 129]. The average 11-12 measurements per day required can be an excessive clinical burden in other ICUs [174, 270, 271], which could lead to protocol non-compliance [152], potentially affecting GC outcomes. Therefore, STAR is extended in clinically validated virtual trials to 1 to 4-, 5-, and 6- hourly treatment intervals, using 1 to 6-hourly stochastic models with the goal of assessing the safety and performance trade-offs at longer intervention intervals within this proven GC approach.

It is hypothesised there will be some loss of tighter control to the narrower, potentially safer 4.4-7.0 mmol/L bands, but lesser loss of performance in the wider, but still safe 4.4-8.0 mmol/L band [76, 77,

79]. Major questions arise over safety from mild and severe hypoglycaemia [46, 163] over longer intervals, and any impact from any resulting reductions in nutrition delivery [29].

In this study, 1 to 3-, 4-, 5-, and 6- hourly versions of STAR are simulated to better capture the effect of increased measurement intervals on STAR GC safety and performance. Stochastic models to predict future SI 1-6 hourly are built from retrospective patient data, using kernel-density methods [124, 127, 256], where SI is identified hourly from clinical data [130, 131]. As this is a first study analysing longer treatment intervals, the well proven 2D stochastic model approach is used.

Five-fold cross validation is used to build new 1-6 hourly stochastic models using 80% of patient data (by patient). The resulting model is then tested using the new extended version of STAR on the other 20% of patient data, where all five test sets are reported in aggregated results. This approach ensures independent development and test sets, and a more robust analysis ensuring stochastic models are not biased by outlying patients or small sub-cohorts.

12.2.2. Virtual Trials

To compare the impact of longer treatment intervals on GC outcomes, validated virtual trials are used to simulate different protocol designs on virtual patients [94]. This approach allows comparison of the safety and performance of the original STAR 1-3 hourly [95, 96, 123], with STAR 1 to 4-, 5-, or 6- hourly on the same underlying virtual patients [91], and is further explained in Chapter 3. In these *in-silico* simulations, virtual patients, including starting BG levels and nutrition rates, were based on initial starting clinical data.

STAR, in its current version, always ensures safety, not allowing the 5th percentile of future BG below the lower limit of the target band (4.4 mmol/L), regardless of the resulting 95th percentile, all of which is a function of the risk-based dosing approach. Because a 3-h measurement interval is relatively short in a clinical sense, the 95th percentile is rarely above 8.5 mmol/L, which is considered acceptable, and nutrition in this case is not decreased. However, as measurement interval increases, wider 5th-95th percentile prediction ranges of BG are more likely, induced by higher potential variability [126, 127, 248, 249], resulting in predicted 95th percentile BG potentially much higher than 8.5 mmol/L.

To mitigate this impact of rising hyperglycaemia over longer intervention intervals, a second version of the protocol is implemented. In this case, the 95th percentile of predicted BG must be lower than 8.5 mmol/L for the treatment intervention to be considered, which is accomplished (where necessary) by reducing nutrition. This approach will decrease the increased risk of hyperglycaemia and show improved efficacy, but could also increase workload and/or reduce nutrition delivery, both of which are clinically desirable "rewards". This second protocol approach is denoted STAR-ULC (STAR Upper Limit Controlled).

The combination of analysing two STAR protocol approaches (STAR and STAR-ULC, Table 12.1) over extended 4-6 hourly intervals limits the analysis and provides the full range of possible performance and safety trade-offs.

Table 12.1 - Summary of STAR and STAR-ULC protocol designs compared in this analysis.

	STAR	**STAR-ULC**
Type	Model-based	Model-based
Target band	4.4-8.0 mmol/L	4.4-8.0 mmol/L
Measurement intervals	1-6 hourly	1-6 hourly
Insulin strategy	Max 9U/h	Max 9U/h
Nutrition strategy	Modulation between 30-100% GF	Modulation between 30-100% GF
Treatment selection strategy (predicted BG)	5th percentile ≥4.4 mmol/L and 95th minimised	5th percentile ≥4.4 mmol/L and 95th percentile ≤8.5 mmol/L

12.2.3. Patient Cohorts

The cohort of patient used is similar to the one used previously in Chapter 9, totalling 681 patient GC episodes ≥ 10 hours and with starting BG > 7.0 mmol/L, from the SPRINT, STAR Christchurch, and STAR Gyula cohorts (Chapter 4). This represents 59439 hours of control.

12.2.4. Comparison Analysis

Safety, efficacy, BG achieved, insulin and nutrition rates, and workload are compared. BG is hourly resampled to allow fair comparison between protocols. Safety is compared using %BG outside target band (%BG < 4.4 mmol/L and %BG > 8.0 mmol/L) and %BG below severe hypoglycaemic threshold (%BG < 2.2mmol/L). Performance is analysed using %BG in the 4.4-8.0 mmol/L target band and median

BG levels achieved. Per-patient insulin (U/h) and nutrition rates (%GF)) are also compared, and workload is assessed using average number of measurements per day.

Additionally, the proportion of patients with ≥ 50% BG in 4.4-7.0 mmol/L and 4.4-8.0 mmol/L are compared for each protocol. High percentage time in these bands, and low incidence of hypoglycaemia, are associated with improved outcomes in ICU patients [10, 11, 32, 38, 50, 76, 77, 79]. Hence, comparing the number of patients reducing / improving time in these bands provides a further outcome-based means to quantify whether patient GC outcomes improved, or not. The number of patients experiencing severe hypoglycaemia is also compared.

The main outcome of the study is to show and evaluate the risk and reward trade-off where:

- Risks are safety (hypoglycaemia), efficacy (performance of GC control), and nutrition provided,
- Reward is the lower workload, reflected by lower measurements per day with the longer treatment intervals used.

This study thus analyses STAR's design robustness as measurement timeframes increases, where, as per protocol design, a reduction in workload (reward) is expected, but at the cost of reduced safety and performance (risks).

12.3. Results

12.3.1. Stochastic Model Comparison

Stochastic models represent the probabilities of changes in SI, as calculated from clinical data. Example 2D stochastic models for predictions 1-6 hours ahead are presented in Figure 12.1, where the 5th and 95th percentiles for future SI at a given current SI are shown. The probability distribution within these bounds would be described by a 3-D 'mountain range' sticking out of the page, approximately centred on the 1-1 line.

Intra-patient variability becomes more similar as prediction interval time increases, and the prediction lines converge to a similar range. This result clearly shows, while a bigger difference can be observed from 1-3h in SI evolution, the difference in intra-patient variability becomes similar when longer intervals

are considered. This outcome can represent a general, conservative, range of intra-patient variability, but alternatively may represent the average of more and less variable patients, which could result in reduced safety in some cases.

More specifically, the longer interval model ranges may “hide” a larger range of changes (rising and falling) before returning to range, increasing the risk of larger unexpected glucose excursions. As a narrower range of possible SI outcomes translates directly to a narrower range of possible BG outcomes for a given treatment. More aggressive dosing can be used for shorter treatment intervals with narrower ranges, as predictions of future SI variability is tighter compared to longer intervals with wider prediction ranges. Thus, the larger the measurement interval, the more conservative the treatment, given the likely higher potential sudden extreme changes in SI.

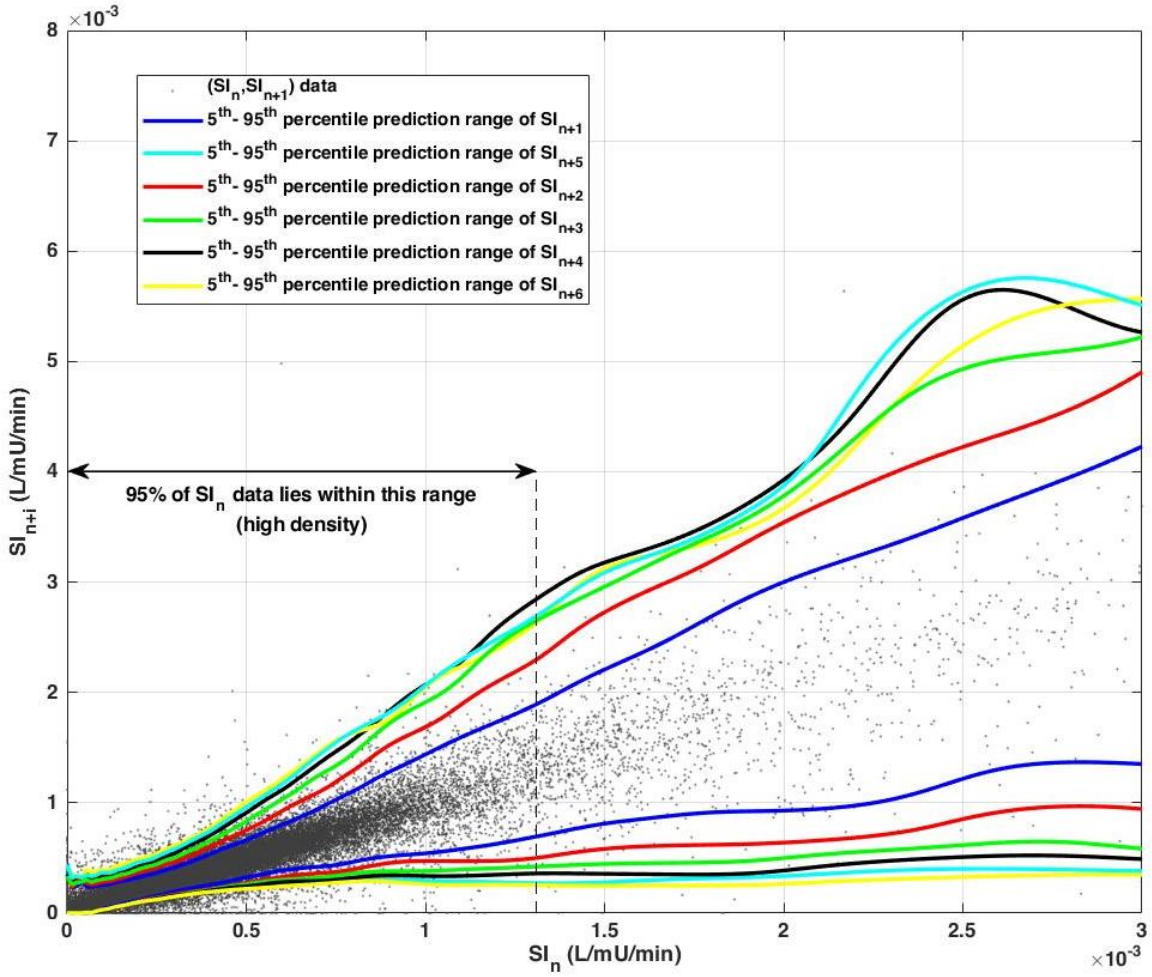

Figure 12.1 – Stochastic model representation showing the 5th-95th percentile prediction range of future 1-6h SI levels given current identified patient-specific SI_n. Data density is higher for lower SI values, explaining the corresponding tighter 5th-95th percentile prediction ranges around the 1-1 line.

12.3.2. STAR Virtual Trial Results

Five-fold cross validation virtual trial results using virtual patients, or 'digital twins' derived from clinical data, are presented in Table 12.2, for each version of STAR (1 to 3-, 4-, 5-, and 6- hourly). These digital twins allow analysis of BG response cohort to different treatment protocols to be compared in both individual patients and the overall cohort. Each arm has the same number of patients, but can have a slightly different number of GC hours, depending on the last measurement interval used in each virtual patient trial (i.e.: if last treatment is 3-hourly vs. 6-hourly, there will be 3 extra simulated hours of GC for this patient). Excerpts from two virtual patient trials comparing STAR-3H and STAR-6H are also presented in Figure 12.2 and Figure 12.3.

As expected, workload decreased as measurement interval increased (from 12 to 8 measurements per day for STAR-3H to STAR-6H). Time in the 4.4-8.0 mmol/L target band was high and similar in all scenarios (80-83%), but with a clear shift upward in median BG levels (6.5 [5.9 7.3] mmol/L for STAR-3H to 6.9 [6.3 7.7] mmol/L for STAR-6H), as reflected in the decreasing % BG in 4.4-7.0 mmol/L. Additionally, the number of patients with ≥ 50% BG in the tighter, safer 4.4-7.0 mmol/L (68% to 55%) and the wider, safe 4.4-8.0 mmol/L (86% vs 84%) slightly decreased, where additional analysis showed 80% of these patients dropping below 50% in those ranges were typically going to higher BG ranges, and 20% where going to lower BG ranges.

Table 12.2 – Virtual trial results of STAR Standard for 1 to 3-, 4-, 5-, and 6- hourly measurements intervals.

	STAR-3H	**STAR-4H**	**STAR-5H**	**STAR-6H**
# GC Episodes	681	681	681	681
# GC hours	59240	59528	59782	60003
# BG meas.	28961	24792	22243	20272
Workload (meas. per day)	12	10	9	8
Median BG (mmol/L)	6.5 [5.9 7.3]	6.7 [6.1 7.5]	6.8 [6.2 7.6]	6.9 [6.3 7.7]
Median Insulin (U/h)	3.2 [2.0 5.0]	3.0 [2.0 4.0]	2.5 [2.0 3.5]	2.5 [1.5 3.0]
Median Nutrition (%GF)	100 [85 100]	95 [80 100]	90 [80 100]	90 [75 100]
%BG in 4.4-8.0 mmol/L	83	82	81	80
%BG in 4.4-7.0 mmol/L	65	59	55	52
%BG >8.0 mmol/L	15	16	17	18
%BG < 4.4 mmol/L	1.6	1.5	1.5	1.6
%BG < 2.2 mmol/L	0.03	0.02	0.04	0.06
# patients ≥50%BG in 4.4-7.0 mmol/L (%)	466 (68%)	432 (63%)	401 (59%)	372 (55%)
# patients ≥50%BG in 4.4-8.0 mmol/L (%)	589 (86%)	583 (86%)	573 (84%)	571 (84%)
# patients min BG < 2.2 mmol/L (%)	14 (2.1%)	12 (1.8%)	18 (2.6%)	19 (2.8%)

Results are based on hourly resampled BG. Median [IQR] is given for per-patient statistics, where appropriate.

Incidence of hyperglycaemia is slightly higher as the interval increased. Most importantly, the incidence of severe hypoglycaemia increased as measurement interval increased, and the number of patients experiencing severe hypoglycaemia also increased (from 14 to 19 patients between STAR-3H and STAR-6H, 2.1% to 2.8% by patient). Interestingly, hypoglycaemia decreased in STAR-4H, with only 12 (1.8%) patients experiencing severe episode.

Overall, these results were achieved with lower insulin and nutrition rates as intervals increased. However, the nutrition rates remained high in these scenarios, where only 25% of patients received less than 75% GF in the worst case (STAR-6H). Thus, there was also some increased hyperglycaemia, as noted.

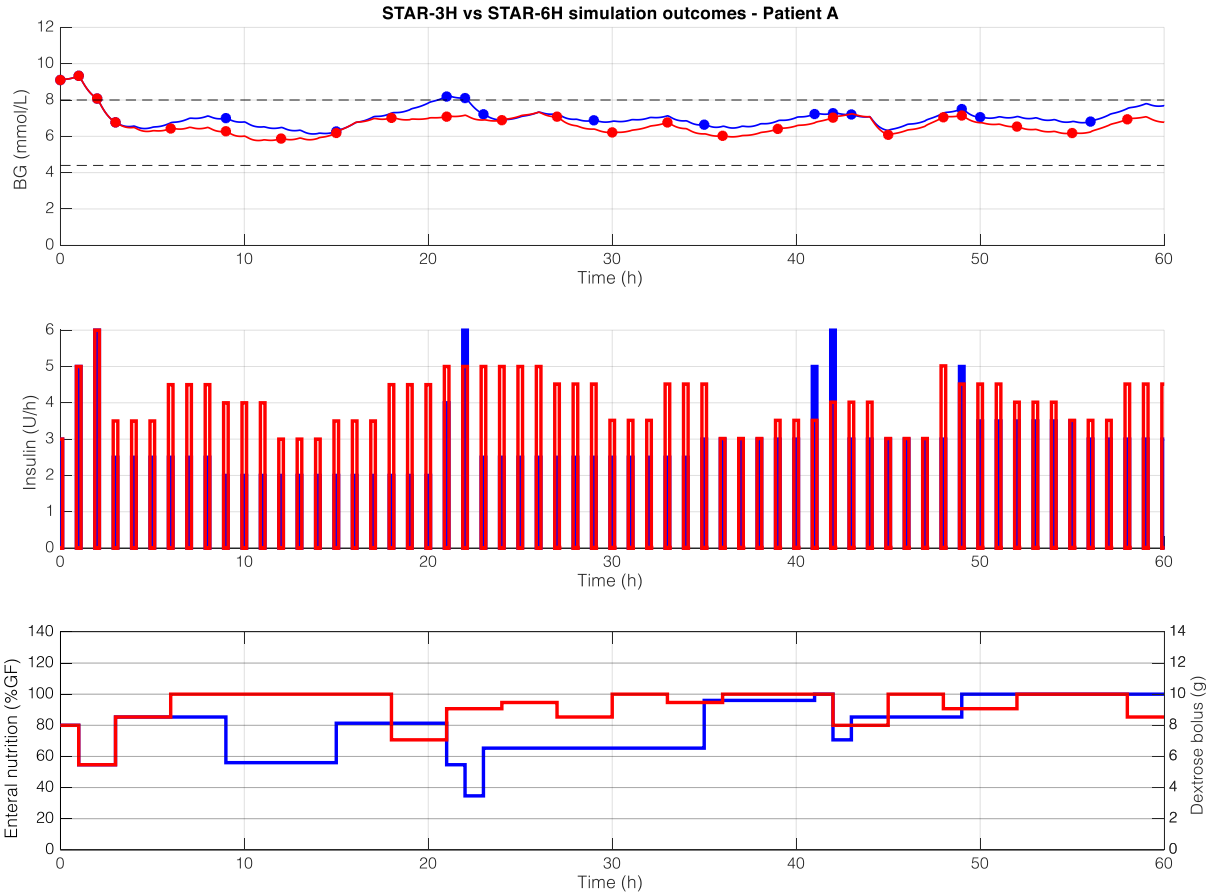

Figure 12.2 – Excerpt of virtual trial results for Patient A. Blood glucose (top), insulin rates (middle), and enteral (solid lines) and dextrose bolus (bars) nutrition rates (bottom) are compared between STAR-3H (red) and STAR-6H (blue).

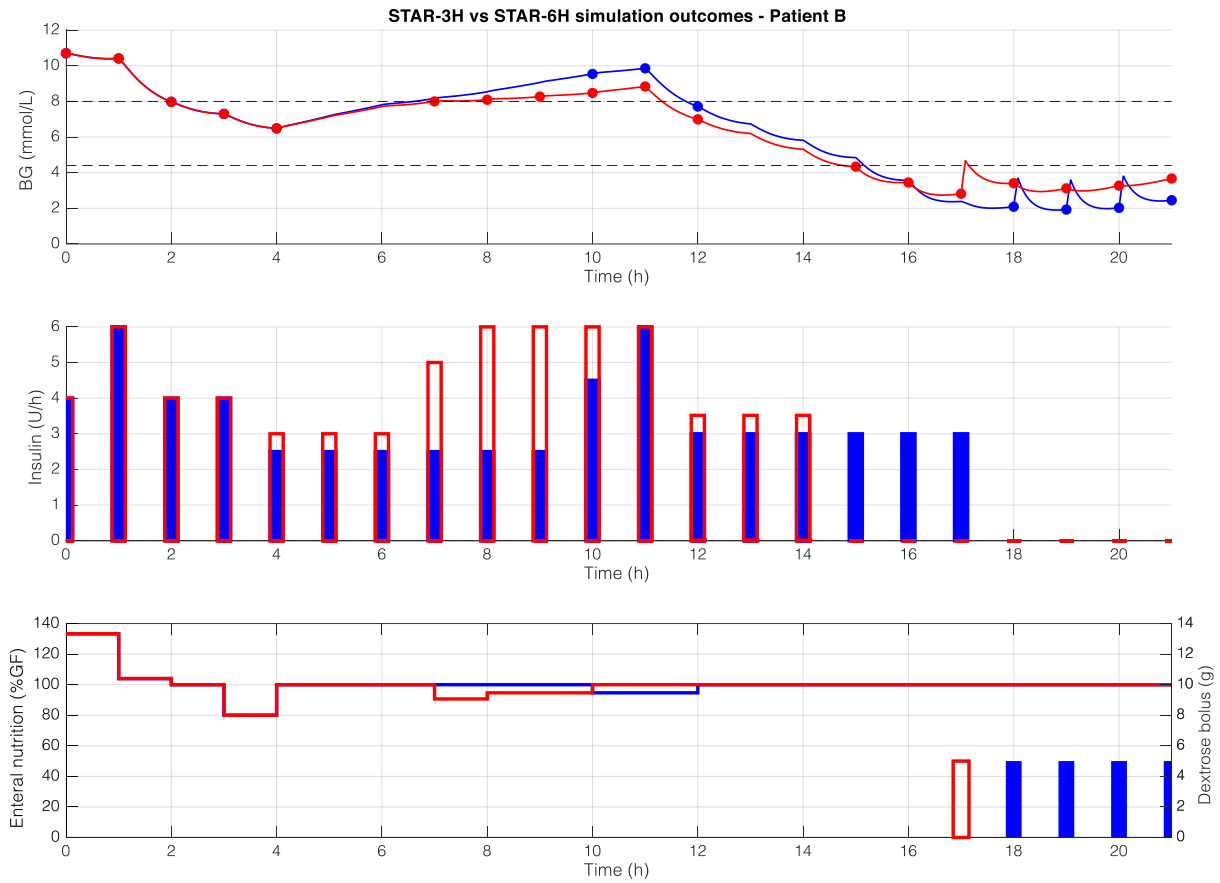

Figure 12.3 – Excerpt of virtual trial results for Patient B. Blood glucose (top), insulin rates (middle), and enteral (solid lines) and dextrose bolus (bars) nutrition rates (bottom) are compared between STAR-3H (red) and STAR-6H (blue).

Table 12.3 – Virtual trial results of STAR-ULC 1 to 3-, 4-, 5-, and 6- hourly, forcing the predicted 95th BG percentile ≤ 8.5 mmol/L.

	STAR-ULC-3H	STAR-ULC-4H	STAR-ULC-5H	STAR-ULC-6H
# GC Episodes	681	681	681	681
# GC hours	59203	59392	59614	59845
# BG meas.	31204	27196	24769	23387
Workload (meas. per day)	13	11	10	9
Median BG (mmol/L)	6.4 [5.9 7.2]	6.5 [6.0 7.3]	6.5 [6.0 7.3]	6.5 [6.0 7.3]
Median Insulin (U/h)	3.0 [2.0 4.5]	2.5 [1.7 4.0]	2.0 [1.5 3.5]	2.0 [1.5 3.5]
Median Nutrition (%GF)	95 [80 100]	75 [65 85]	70 [60 80]	60 [50 75]
%BG in 4.4-8.0 mmol/L	84	84	85	85
%BG in 4.4-7.0 mmol/L	68	67	67	67
%BG >8.0 mmol/L	14	14	14	14
%BG < 4.4 mmol/L	1.6	1.5	1.5	1.5
%BG < 2.2 mmol/L	0.02	0.02	0.02	0.04
# patients ≥50%BG in 4.4-7.0 mmol/L (%)	497 (73%)	486 (71%)	481 (71%)	481 (71%)
# patients ≥50%BG in 4.4-8.0 mmol/L (%)	597 (88%)	596 (88%)	594 (87%)	592 (87%)
# patients min BG < 2.2 mmol/L (%)	11 (1.6%)	9 (1.3%)	9 (1.3%)	15 (2.2%)

Results are based on hourly resampled BG. Median [IQR] is given for per-patient statistics, where appropriate.

12.3.3. STAR Upper Limit Controlled (STAR-ULC) Virtual Trial Results

An 'Upper Limit Controlled' approach is also analysed, in which nutrition is modulated so the upper 95th percentile of possible BG outcomes does not exceed 8.5 mmol/L. This approach reduces hyperglycaemia, as well as the increased risk associated with large insulin and nutrition doses, which amplifies uncertainty in SI, especially as the measurement interval increases. Five-fold cross validation results of the 1 to 3-, 4-, 5-, and 6- hourly versions of the STAR Upper Limit Controlled (STAR-ULC) approach, forcing the 95th percentile of BG ≤ 8.5 mmol/L are presented in Table 12.3.

High performance (~84% in target band and ~67% in 4.4-7.0 mmol/L) and high safety (14% BG > 8.0 mmol/L and 1.5% BG < 4.4 mmol/L) were achieved, and this result was surprisingly very similar regardless of measurement intervals. The number of patients experiencing severe hypoglycaemia decreased compared to STAR Standard (Table 12.2). STAR-ULC-4H (9 patients) and STAR-ULC-5H (9 patients) had both reduced number of patients experiencing hypoglycaemia compared to STAR-4H (12 patients) and STAR-5H (18 patients). These values were also lower compared to STAR-ULC-3H (11 patients) and STAR-ULC-6H (15 patients). This result reflects a reduction in risk due to reduced insulin dose by limiting the upper glycaemic as well within the STAR risk-based dosing system.

The number of patients with ≥ 50% BG in 4.4-7.0 mmol/L (~71%) and 4.4-8.0 mmol/L (~87%) was similar across all measurement intervals, reflecting effective control was achieved consistently for most patients. These numbers are higher compared to STAR Standard (Table 12.2), especially when comparing the tighter, safer 4.4-7.0 mmol/L band (55-68% for STAR Standard vs. 71-73% for STAR-ULC), which would reflect a significant improvement in outcomes [77, 78].

Improved safety and efficacy were achieved here with significantly lower insulin and nutrition rates administered (Table 12.3) compared to STAR Standard (Table 12.2). A comparison of STAR-6H and STAR-ULC-6H is presented in Figure 12.4, where this difference is clearly illustrated. Finally, workload increased by 1 additional measurement per day for each version compared to STAR Standard (Table 12.2), but are still lower than STAR 3-h standard of 12 per day [87] at the 4-6 hourly intervals with better performance and safety.

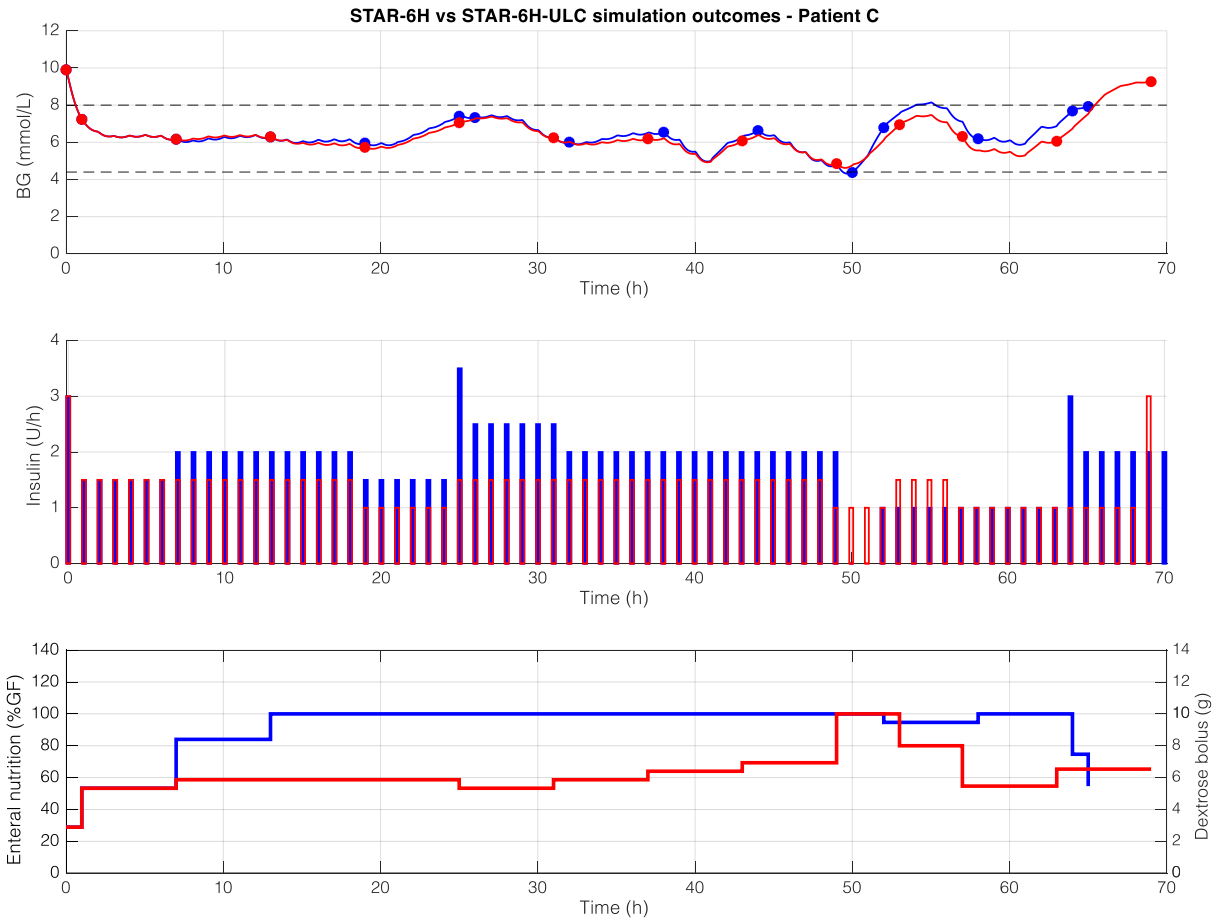

Figure 12.4 – Excerpt of virtual trial results for Patient C. Blood glucose (top), insulin rates (middle), and enteral (solid lines) and dextrose bolus (bars) nutrition rates (bottom) are compared between STAR-6H (blue) and STAR-ULC-6H (red).

12.4. Discussion

Foremost, it is important to understand metabolic variability, reflected in inter- and intra- patient variability, is what makes GC hard to achieve safely [86, 135]. Therefore, it is critical for GC protocol design to account for both, using dynamic, personalised solutions [98]. While the use of physiological models allows direct identification of inter-patient variability [99], STAR is the only current protocol also using stochastic modelling to evaluate intra-patient variability [95], which it then employs in a unique risk-based dosing strategy.

In Chapter 6 comparing survivors and non-survivors, inter-patient variability has been shown different while intra-patient variability was clinically equivalent [135]. Therefore, this result emphasizes the importance of identifying key physiological parameters, such as SI here, and assessing potential variability to provide safe, and effective control for all, which is critical to improving outcomes [38, 160].

In addition, compliance to protocol is essential to ensure any clinical judgement bias in results outcomes and conclusions [160], where longer intervals may improve compliance [152].

In many ICUs, protocols are often 4-hourly based once BG levels are stabilised in the target band. In practice, this interval can quickly become 5- or 6- hourly, given clinical judgement and excessive clinical workload [8, 85, 86, 160, 268]. Usually, the higher the target band, the greater the permissive hyperglycaemia, and, indirectly, the lower the risk of hypoglycaemia. However, it is also important to keep in mind it is impossible to clearly know whether the patient suffered from hypoglycaemia over longer measurement intervals without CGM or similar [272, 273].

This study assesses the potential to reduce workload with the safe, and effective STAR GC framework, and the impact on safety and performance. Specifically, extreme changes in SI levels between consecutive measurements have a greater chance to occur as measurement intervals increase. Typically, for a given insulin dose and a sudden rise in SI, BG levels will suddenly drop. If this drop occurs one hour after treatment intervention and the next measurement is due in 5 hours, it can have significant impact on patient BG, seen in the increased number of patients experiencing severe hypoglycaemia (Table 12.2). However, when limiting GC to lower measurement intervals, this sudden reduction in BG levels will potentially be seen sooner, and treatment adapted, potentially averting severe hypoglycaemia. This scenario is shown in Figure 12.3, where Patient B becomes more insulin sensitive at 12h, and where STAR-3H captures this behaviour and can adapt treatment faster (at 15h) compared to STAR-6H where severe hypoglycaemia occurs (at 18h).

As seen in Figure 12.2 - Figure 12.4, the different GC scenarios, based on the different measurements intervals allowed, led to significant different measurement timing. Therefore, while one version could by chance measure BG right before hypoglycaemia, another could fail due to unfortunate timing. This issue adds difficulty when interpreting results, but reflects real practice, where measurement timing is also a factor influencing control. In clinical practice, despite nurse selecting a specific treatment interval, the new measurement may be taken a few minutes or even hours later, or earlier. This measurement timing may thus (unexpectedly) influence results, as seen in Table 12.3, where incidence of severe hypoglycaemia is actually lower for STAR-ULC-4H and STAR-ULC-5H compared to STAR-ULC-3H. However, while this issue is typical in medical environment and time dependant decision making, a large

cohort of virtual patients enables a balanced analysis of the potential advantage/disadvantage. More specifically, the differences reported in Table 12.3 are small, and may thus be considered as 'noise'.

Another potential consequence of increased measurement interval is the lower confidence in future evolution of SI. The 5^{th}-95^{th} percentile of predicted SI is thus wider, and STAR will consistently be more conservative in insulin dosing [127], typically providing lower insulin rates to ensure safety (Table 12.2). While it is a safe approach, performance is affected due to the higher predicted risk, increasing BG levels (Table 12.2). The other effect is a general increase / shift in BG outcomes achieved, leading to higher %BG > 8.0 mmol/L and %BG > 10.0 mmol/L, the severe hyperglycaemic threshold. Interestingly, this outcome is achieved with slightly lower, but still high, nutrition rates [134, 274] to avoid potentially more important hyperglycaemic risk (Table 12.2).

Hence, to reduce the related expected increased hyperglycaemia, an adapted approach forcing the 95^{th} percentile of predicted BG ≤ 8.5 mmol/L, the STAR-ULC approach, was undertaken. Figure 12.5 presents the main risk and reward summary outcome comparison between STAR Standard and STAR-ULC as a function of measurement intervals. Significantly more consistent GC outcomes were achieved regardless of measurement timeframe (Table 12.3, Figure 12.5). Surprisingly, these results show improved safety, and lowered the number of patients experiencing hypoglycaemia. This result and consistency can be explained by the increased workload, increasing the chances to react faster to reduced BG. However, it is most likely due to treatments suggesting lower insulin and nutrition rates (Table 12.3, Figure 12.5), where insulin's impact on BG reduction from a sudden rise in SI was reduced thanks to lower insulin concentrations and concomitantly reduced nutrition.

Virtual patient trials using the STAR-ULC to mitigate the risk of hyperglycaemia due to larger predicted variability resulted in trade-off between BG outcomes, workload, and nutrition rates achieved. Nutrition management in ICU is a hot topic [27-29, 254], where no clear uniform guidelines exist. Recent reviews suggest stepping increased nutrition rates from ICU admission, starting at 25% GF and ideally increasing by 25% every two days to reach 100% within a week [27]. In these results, nutrition rates achieved (60 [50 75] %GF in the worst STAR-ULC-6H case) are still comparable to, or better than, the recommendations, and thus potentially acceptable.

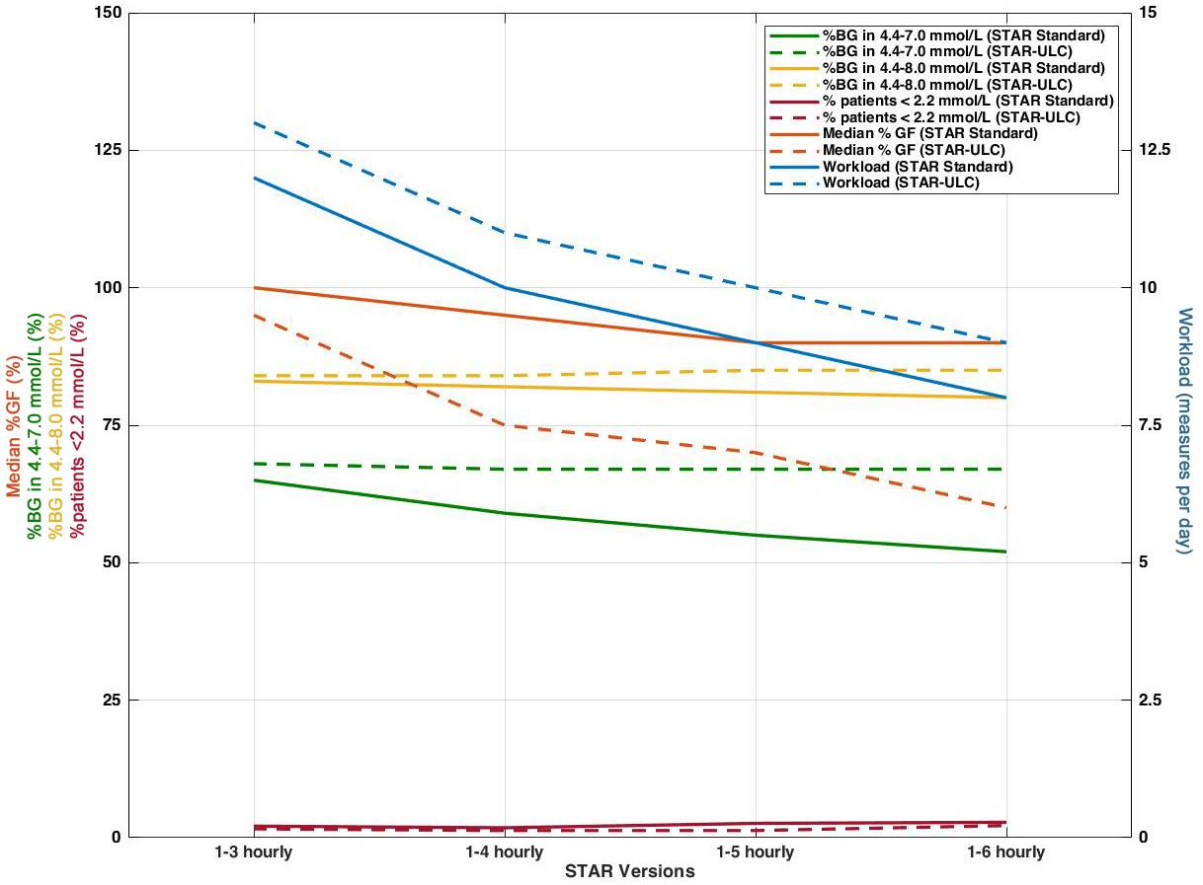

Figure 12.5 - Risk and reward trade-off between STAR Standard (solid) and STAR-ULC (dashed) as a function of intervention intervals.

In addition, these rates achieved with STAR-ULC were comparable to the SPRINT protocol results, which was the only study to reduce all three of mortality, organ failure, and hypoglycaemia [38, 81]. Previous studies showed STAR using 1-3 hour intervals provides close to the best nutrition delivery rates in the world [134] due to its ability to provide personalised nutrition, adapted to patient needs, while always ensuring safety. Hence, these results show the STAR Standard and STAR-ULC approaches can deliver acceptable, but different nutrition delivery rates with extended intervals and reduced workload, presenting a clear trade-off choice.

Ideally, 1-hourly measurements would provide the best outcomes. However, this approach is not clinically feasible and would require too much workload. CGM could also potentially provide improved control [273, 275]. In general, this technology is still not fully reliable in ICUs [191], but may develop further in future to full effect and enable far more flexible control approaches [276].

Overall, the virtual trial results are encouraging, and, regardless of measurement interval, provided safe and effective control for nearly all patients. Consistent high %BG in the tighter, safer 4.4-7.0 mmol/L and

wider, still safe 4.4-8.0 mmol/L target band are associated with improved outcomes in ICUs [38, 76, 77]. Results suggest STAR is robust when using longer treatment intervals, and can safely adapt treatment to patient needs. However, these results also show the inevitable risk and reward trade-off between measurement interval and GC safety and efficacy. Increasing measurement intervals modestly increases risk of hypoglycaemia from 1.6% of patients to 2.2% or 2.8% (Tables 1-2), which are still very low compared to many prior studies [65-67]. The potentially bigger trade-offs come between nutrition delivery and desired performance, both compared to workload.

More specifically, reducing workload using longer treatment intervals results in slightly high incidence of hyper- and hypo- glycaemia, given higher potential future SI variability. STAR-ULC provides safer, more effective, and tighter control compared to STAR Standard, at the cost of slightly increased workload and lower nutrition and insulin rates. This outcome suggests high nutrition and insulin rates magnify uncertainty as treatment interval increases, which should be expected. Reducing nutrition (and thus insulin) thus reduces risk of hypoglycaemia, further emphasising this “workload-performance-nutrition” risk and reward trade-off. While 4-hourly measurements are common in GC, whether 5- and 6- hourly are suitable in clinical practice is an important question.

The only major change in the STAR GC protocol design in this analysis is the ability to suggest longer treatment intervals, given these treatments meet safety requirements, using additional corresponding extended stochastic models. Nothing else was changed from the original protocol. However, further analysis could consider some kind of hybrid system, with more restriction for longer treatment intervals (such as a potential reduced upper limit of insulin rate), to avoid additional risks. While this change could be considered, results presented here still show very high safety compared to most published protocols [8, 65, 66], and, thus, such changes to the original protocol seem less necessary.

The results presented here use virtual patient and trial simulations [128]. Such simulations use a physiological model, where some physiological parameters are approximated, and, thus, could potentially lead to some minimal bias [93]. However, the model used has been validated and extensively clinically used in a wide range of clinical scenarios [87, 91, 94, 128, 136, 137, 150, 154, 259]. It is also proven to reflect what is seen clinically by accurately predicting subsequent clinical results [96, 105]. However, virtual trials represent ideal conditions, with full compliance to protocol. Results may thus be

a best case compared to reality, but representative of the reality and generalisable to other population cohort. Hence, all results presented should be validated in future clinical pilot trials, which are justified by the results presented here.

12.5. Summary

In this study, the STAR GC framework is shown to provide safe, effective control to nearly all patients, despite increasing measurement intervals from 3- to 6- hourly to reduce workload. However, longer treatment intervals are associated with modestly increased risks of hyper- and hypo- glycaemia, as well as potential reductions in nutrition delivery when these risks are mitigated by limiting hyperglycaemic risk. The overall results present a clear risk and reward trade-off between workload and GC outcomes within the context of this proven risk-based GC framework. Overall, STAR's unique risk-based dosing approach is robust to adaptation to using longer treatment intervals. Clinical pilot trials using STAR with different measurement timeframes should be undertaken to confirm these results clinically.

Chapter 13: Conclusions

Stress-induced hyperglycaemia is common in critically ill patients, leading to excessive glucose production and increased insulin resistance. The resulting excessive BG levels are associated with increased morbidity and mortality. Tight GC has shown beneficial impact, but is hard to achieve safely and effectively due to high levels of inter- and intra- patient variability. However, the increased risk of hypoglycaemia with GC, and its independent association with increased mortality, has been identified has a potential safety barrier for GC targeting normoglycaemic ranges. Hence, there is ongoing debate on the optimal glycaemic targets, considering possible benefits against the consequences of the widely shown increased hypoglycaemic risks with lower target bands.

However, causality of the association of hypoglycaemia with tight GC remains unclear. This thesis first aimed to better understand what makes achieving (safe) GC hard. It then examines whether safe GC can be achieved for all patients, suggests solutions to achieve precision GC in an ICU context, and, finally, addresses major issues impacting the future of GC in the ICU.

This chapter presents the main conclusions of this thesis, by addressing the four questions raised in Chapter 1.

13.1. The Keys to GC Success

What are the main factors influencing high quality GC, and why did some studies successfully provide safe, effective control, while others did not?

Most guidelines, to date, recommend higher targets based on studies failing to provide safe control for all patients, as a “first do no harm” approach. However, many other studies have demonstrated the ability to achieve safe, tight GC with patient-specific model-based methods. Virtual trial analysis of two contradicting studies (NICE-SUGAR and STAR), on the same underlying virtual cohort, enabled a better understanding of the issues and identified reasons for NICE-SUGAR’s increased incidence of hypoglycaemia, despite targeting normoglycaemic ranges.

Virtual trial results provided evidence that poor compliance to protocol can bias results, thus questioning the conclusions of the NICE-SUGAR, with implications for the interpretation of similar studies. More specifically, simulations of the protocol showed significant differences with reported clinical results, despite providing relatively safe and tight control. The excessive BG measurements required by NICE-SUGAR per protocol (~1-hourly) was clearly a design failure, as they were not achieved in practice. In addition, the absence of specific limits on the insulin dosing may have led to further non-compliance, and thus further biased safety and performance outcomes.

Virtual trial results also showed NICE-SUGAR’s lack of patient-specificity resulted in lower GC safety and efficacy. In contrast, STAR’s unique patient-specific risk-based dosing approach significantly decreased hypoglycaemia and provided tighter control. Accounting for inter- and intra- patient variability is thus also a key to GC success.

These results bring new important considerations when assessing GC clinical trial outcomes, suggesting assessing how well GC is implemented is an absolute requirement before assessing clinical outcomes. Importantly, the results suggest hypoglycaemia has been wrongly blamed for poor patient outcomes, which are instead due to the ineffectiveness of GC design.

13.2. Safe, Effective GC for All

Is poor GC due to patient severity and outcome, and thus unavoidable? Or should everyone be able to benefit from equally safe and effective control, regardless of clinical outcome?

While protocol design is a key to GC success, the associated increased hypoglycaemia in some studies could be considered as a consequence of underlying patient state, rather than poor GC protocol design. In turn, this question would suggest more severely ill patients would benefit less from GC, and, thus, there would be a lesser necessity to control glycaemia beyond a modest lowering for at least these patients.

In a retrospective analysis comparing survivors and non-survivors, SI was shown to be different between the groups, while SI variability was equivalent. Given the equal variability, survivors and non-survivors are equally difficult to control, suggesting similar GC outcomes, safety and performance, should be achieved regardless of patient severity and (eventual) patient clinical outcome. GC outcome is thus a function of GC design, not patient condition, and all patients should be able to benefit from equally safe, and effective GC.

Similar analyses on SI levels and variability for males and females provided the same conclusions. Overall, it suggests while inter-patient variability may be different across patients, intra-patient variability is always equivalent. Thus, patients may need different insulin dosage based on their patient-specific SI, but the associated risks are always equivalent. This outcome further establishes the necessity of directly accounting for intra-patient variability.

Hence, failing to provide safe control for nearly all patients can bias clinical trial results, which thus may not reflect metabolic response to treatment, but instead reflect the poor quality of control applied. Given a patient-specific, model-based, and risk-based GC protocol, safe and effective GC should be achieved for all.

13.3. The Leading Role of Metabolic Variability in Precision GC

How can precision GC be achieved for all? And can this patient-specific precision be increased?

Quantifying and directly accounting for inter- and intra- patient variability is essential to provide safe, and effective GC. The STAR GC framework accounts for both. STAR characterised patient-specific SI, accounting for inter-patient variability, and assesses the risks associated with a given treatment using a stochastic model, accounting for intra-patient variability. STAR has been shown to provide safe, effective control for nearly all patients, with high clinical staff compliance to protocol.

While intra-patient variability is similar across patients, it must be clearly characterised to avoid any potential hypoglycaemic risk. The more accurate the predictions of potential sudden change in SI levels are, the better STAR can mitigate the risks of hypoglycaemia in treatment recommendation. A new stochastic model was thus developed accounting for prior changes in SI levels to predict future variability. This new 3D stochastic model better characterises patient-specific metabolic variability, providing much tighter, personalised prediction ranges, and thus better performance with equal or better safety.

This new 3D stochastic model was implemented in a pilot clinical trial in Christchurch, New Zealand, improving both safety and efficacy of STAR, while also providing higher nutrition to patients. It also further validated STAR's ability to provide safe control for all, despite targeting lower, normoglycaemic ranges.

13.4. Hypoglycaemia, Nutrition, and Treatment Intervals: The Risk and Reward Trade-off

What is the risk and reward of longer treatment intervals?

STAR is the only GC protocol also modulating nutrition to achieve safe, and effective GC. A clinical trial was implemented at the University Hospital Centre of Liège, Belgium, to assess STAR's ability to adapt to different ICU practices, but also quantify, for the first time, the impact of also modulating nutrition.

Clinical trial results provided evidence STAR was able to provide significantly safer, more effective GC than the insulin-only version. The overall nutrition rates achieved were more consistently managed with STAR than clinically, and, in theory, much closer to the patient-specific energy expenditure. Modulating nutrition also significantly reduced workload, as indicated by the lower measures required. Compared to the local standard protocol, STAR provided significantly tighter control, and resulted in lower incidence of both hypo- and hyper- glycaemia, proving STAR's ability to generalise across different patient populations and ICU practices.

These encouraging results were achieved with higher workload compared to local standards. This outcome suggested assessing the impact of using longer treatment intervals on GC. Virtual trials were thus simulated to assess the risk and reward trade-off associated with longer treatment intervals in the context of STAR. More specifically, STAR treatment intervals were extended from 1 to 3-hourly to 1 to 4-, 5-, and 6- hourly. While the reward of using longer treatment intervals is the lower workload associated, the risks include lower GC safety and efficacy at similar nutrition levels, or equally safe, effective control with reduced nutrition. This outcome range of compromises allows clinical staff to significantly reduce workload, which is critical to more regular uptake of GC. This analysis is the first time such trade-offs have been quantified, despite the impact of workload and non-compliance on safety and performance.

13.5. Take Home Message

Safe and effective GC to normoglycaemic ranges can be achieved for all patients, regardless of patient condition. This can be done using patient-specific, clinically feasible, GC design accounting for inter- and intra- patient variability, ultimately improving patient outcome. Failing to do so suggests critical GC design failure. It is thus time to rethink GC guidelines, and to undertake clinical trials assessing the impact of GC on clinical outcome, without the bias of avoidable hypoglycaemia.

Chapter 14: Future Work

This thesis provided significant contributions for improved personalised precision GC in the ICU. This chapter suggests potential future work that could be undertaken to further address and characterise the keys for tight GC success.

14.1. Inter- and Intra- Patient Variability

Intra-patient variability, equivalent across patients, centres, and a range of sub-cohorts, has been better characterised in Chapter 9. The improved predictions of future variability thus enabled greater personalisation of GC. However, more work could be done to further improve these predictions. Recent published studies using BG and SI as inputs to predict future variability also showed improved associated GC outcomes [256, 257]. Using even more inputs to predict future SI levels could potentially further improve predictions. The main limitations would be the potential lack of clinical data and data density at this time. Additionally, given the extremely high safety already achieved in STAR, the net gain in GC outcome might not be significant compared to the effort (computational complexity).

Inter-patient variability has been shown to be not equivalent and varies significantly between centres, patients, and sub-cohorts. The patient-specific, model-based identified SI allows STAR to account for inter-patient variability in every intervention. Hence, improving the accuracy of identified SI across patients can also bring significant improvement in the context of STAR. A known problem in the current ICING model used in STAR is the impact of under-estimated EGP production when SI saturates at very low values [277, 278]. A better, more patient-specific estimate of EGP would enable more accurate SI values. However, EFP is not uniquely identifiable with current clinical data This issue has not been addressed in this thesis, and could be analysed in future work.

14.2. Multi-Centre Clinical Trial Assessing Clinical Outcome

It is time to implement new RCTs truly assessing the GC impact on clinical outcome, based on the evidences presented in this thesis. More specifically, given a feasible patient-specific GC protocol, accounting for both inter- and intra- patient variability, all patients should benefit from equally safe, and effective GC. If strict compliance to protocol is achieved across different participating ICUs, and no significant difference in the incidence of hypoglycaemia between patients is observed, then assessing the true impact of GC targeting different ranges will truly be quantified, and it will not be biased by poor protocol design or compliance. In turn, these future RCTs could finally make a case for the implementation of tight, or conventional, GC in all ICUs.

14.3. Clinical Trial of Extended 1-6 Hourly Measurement Intervals

The risk and reward trade-off associated with longer measurement intervals needs to be confirmed clinically. The encouraging results obtained in Chapter 12 suggest the implementation of a pilot trial to confirm these findings and better quantify the associated risks. However, while 4-hourly measurements are common in GC practices, whether 5- or 6- hourly measurement intervals should be used is an important debate to discuss prior clinical implementation. More specifically, longer treatment intervals could lead to important increased risk of hypoglycaemia, without compromises regarding nutrition delivery. This trial would thus open significant consideration of the necessary level of nutrition delivery in critical care.

14.4. Nutrition During Critical Illness

One of the advantages of STAR is its ability to also control nutrition to achieve safer GC. Despite modulating nutrition, STAR has been previously shown to provide close to best nutrition expenditure in the world. The amount of nutrition to provide, the administration route, and the timing are all challenging questions in ICU patients, and vary across patients and over time. While no gold standard exists, current guidelines suggest progressive adaptive nutrition based on days from ICU admission [27, 70], but once again lack patient-specificity.

Recent studies showed early feeding and overfeeding during the acute phase do not improve outcomes compared to trophic or permissive underfeeding, and may have negative effects [28, 274, 279, 280]. Additionally, early parenteral nutrition during the acute phase suppresses autophagy, a critical repair process for critical illness organ failure [281, 282]. However, during the recovery phase, increased nutritional intake is necessary in the transition from catabolism to anabolism to maximize anabolic repair in recovery [29]. Lacking a biomarker capturing metabolic state and its transition, it is impossible to know when and how much nutrition to deliver.

The main challenge is to optimise nutrition based on the patient-specific transition from the acute phase to the recovery phase. Future work should investigate whether identified SI is an accurate, personalised, real-time marker to capture the transition from catabolic to anabolic metabolism, reflecting patient-specific ability to utilise glucose. In turn, this model-based identified SI parameter could be used to

identify the transition between acute and recovery phases, and improve nutrition intake for critically ill patients.

Appendix I: Metabolic System and Insulin Sensitivity

This Additional File is designed to present the model and methods used in several referenced studies (e.g. [81, 86, 92, 97, 105, 106, 129, 283, 284]) in this paper. The presentation is brief, relying on a separate set of references (from the main article) given at the end of this Appendix, which interested readers can use for explicit details on any aspect of this model and the methods used.

A clinically validated computer model of the metabolic system [94] was used to identify [131] patient-specific, time-varying (hourly) insulin sensitivity (SI) every hour. The model presented is a compartment model, accounting for the appearance of insulin and glucose in blood and interstitial fluid volumes. Figure A1-1 shows this model schematically.

$$\dot{G}(t) = -p_G G(t) - S_I G(t)\frac{Q(t)}{1+\alpha_G Q(t)} + \frac{P(t) + EGP - CNS}{V_G} \quad \text{A1.1}$$

$$\dot{Q}(t) = n_I\big(I(t) - Q(t)\big) - n_C\frac{Q(t)}{1+\alpha_G Q(t)} \quad \text{A1.2}$$

$$\dot{I}(t) = n_K I(t) - n_L\frac{I(t)}{1+\alpha_I I(t)} - n_I\big(I(t) - Q(t)\big) + \frac{u_{ex}(t)}{V_I} + (1-x_L)\frac{u_{en}(G)}{V_I} \quad \text{A1.3}$$

$$P(t) = \min(d_2 P2,\ P_{\max}) + PN(t) \quad \text{A1.4}$$

$$\dot{P1}(t) = -d_1 P1 + D(t) \quad \text{A1.5}$$

$$\dot{P2}(t) = -\min(d_2 P2,\ P_{\max}) + d_1 P1 \quad \text{A1.6}$$

$$u_{en}(G) = \min\left(\max\left(u_{min}, k_1 G(t) + k_2\right), u_{max}\right) \quad \text{A1.7}$$

Where G(t) [mmol/L] is plasma glucose concentration, I(t) and Q(t) [mU/L] are plasma and interstitial insulin concentrations. Pancreatic insulin secretion is modelled as a function of plasma glucose and is denoted $u_{en}(G)$. The associated parameter values and descriptions are listed in Table A1-1. Table A1-2 shows the exogenous input variables to the model.

Table A1-1. Parameter values and descriptions for the glucose-insulin model.

	Value	Description	Fixed?	Identification Method	Reported range
$S_I(t)$	l/mU/min	Insulin sensitivity	N	Integral based fitting [131]	-
α_G	1/65 (0.015) l/mU	Saturation of insulin-mediated glucose uptake	Y	Chosen from literature [97]. Sensitivity tested in [141].	0.001 – 0.025 l/min [285].
p_G	0.006 min^{-1}	Other non-insulin mediated glucose clearance	Y	Identified: grid search and error minimisation [93]	0.004 – 0.047 min^{-1} [93]
V_G	13.3 L	Glucose distribution volume	Y	Chosen from literature [93]	10.0 – 15.75 L [140] 0.22 L/kg [286]
EGP	1.16 mmol/min	Endogenous glucose production (hepatic)	Y	Grid search and error minimisation [93]. Later (unsuccessful) analysis as function of glucose and time [82].	0.10 – 2.36 mmol/min [82].
CNS	0.3 mmol/min	Glucose uptake by central nervous system	Y	Chosen from literature [93].	0.29 – 0.38 mmol/min [93].
x_L	0.67	Fractional first pass hepatics insulin clearance from portal vein	Y	Chosen from literature [93]	0.5-0.95 [138].
n_L	0.1578 min^{-1}	Rate parameter: general hepatic insulin clearance	Y	Chosen based on previous work [93]	0.1 – 0.21 min^{-1} [140]
α_I	1.7x10^{-3} l/mU	Saturation of hepatics insulin clearance	Y	Chosen from literature [285].	0.0005 – 0.0043 L/mU [285].
n_K	0.0542 min^{-1}	Rate parameter: kidney clearance of insulin	Y	Chosen from literature [93].	0.053–0.064 min^{-1}[140].
n_C	0.006 min^{-1}	Rate parameter: cellular degradation of internalised insulin	Y	Identified: grid search and error minimisation [141]	Parameter sensitivity: [141].
n_I	0.006 min^{-1}	Rate parameter: diffusion of insulin between plasma and interstitium	Y	Identified: grid search and error minimisation [141]	0 – 0.06 min^{-1} [141].
k_1	14.9 mU·l/mmol/min	Insulin secretion model parameter	Y	Model fit to clinical C-peptide and Insulin data. Compared to results derived from literature [82].	8 - 45.9 mU/min [82].
k_2	-49.9 mU/min	Insulin secretion model parameter	Y	Model fit to clinical C-peptide and Insulin data [82].	-
u_{min}	16.7 mU/min	Minimum insulin secretion	Y	Constraint derived from lower range of clinical insulin secretion data [82].	-
u_{max}	266.7 mU/min	Maximum insulin secretion	Y	Constraint derived from upper range of clinical insulin secretion data [82].	-
V_I	4.0 L	Insulin distribution volume	Y	Chosen from literature [82].	3.15 – 4.75 L

Table A1-2. Exogenous input variables to the glucose-insulin model.

Variable	Unit	Description
PN(t)	mmol/min	Intravenous glucose input rate (parenteral nutrition)
D(t)	mmol/min	Oral glucose input rate (enteral nutrition)
$u_{ex}(t)$	mU/min	Intravenous insulin input rate

The insulin sensitivity SI can be identified hourly from blood glucose data along with the clinical insulin and nutritional inputs from all sources [131, 287]. SI is also the critical parameter in predicting the outcome of a nutrition and/or insulin intervention in this model, based on the definition above [97, 131, 283]. It represents the whole body balance of insulin and CHO from all sources. SI can vary with patient-status hour to hour, with larger acute changes or smaller gradual evolution. Thus, the identified SI can be used to characterise metabolic response and evolution for cohorts or specific-patients, enabling more optimal and robust dosing [124, 125, 157, 185]. Two example SI profiles and model fit to clinical data can be found in Figure A1-2. Both show stable BG within the 4.4 – 8.0 mmol/L range, despite different underlying insulin sensitivity variability and the insulin and nutrition doses required to achieve comparable BG stability.

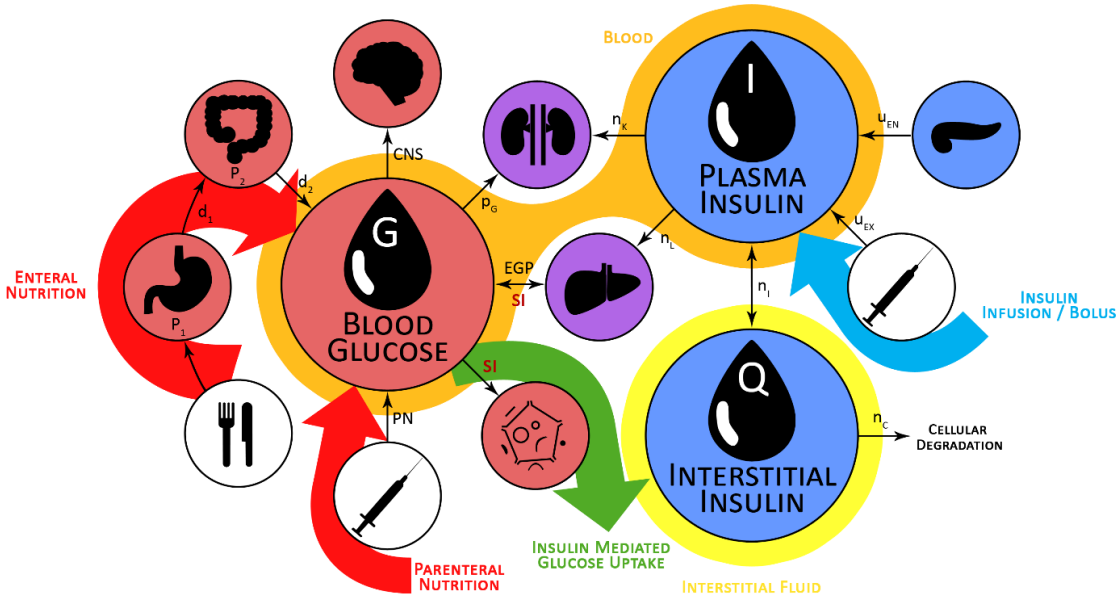

Figure AI-1: Model schematic for Equations (A1.1)-(A1.3) showing the physiological compartments and clearances, as well as the appearance of exogenous insulin and carbohydrate, and their kinetic pathways. Insulin sensitivity (SI) can vary over time (hour to hour) thus affecting glycaemic outcomes for a given insulin and/or nutrition intervention.

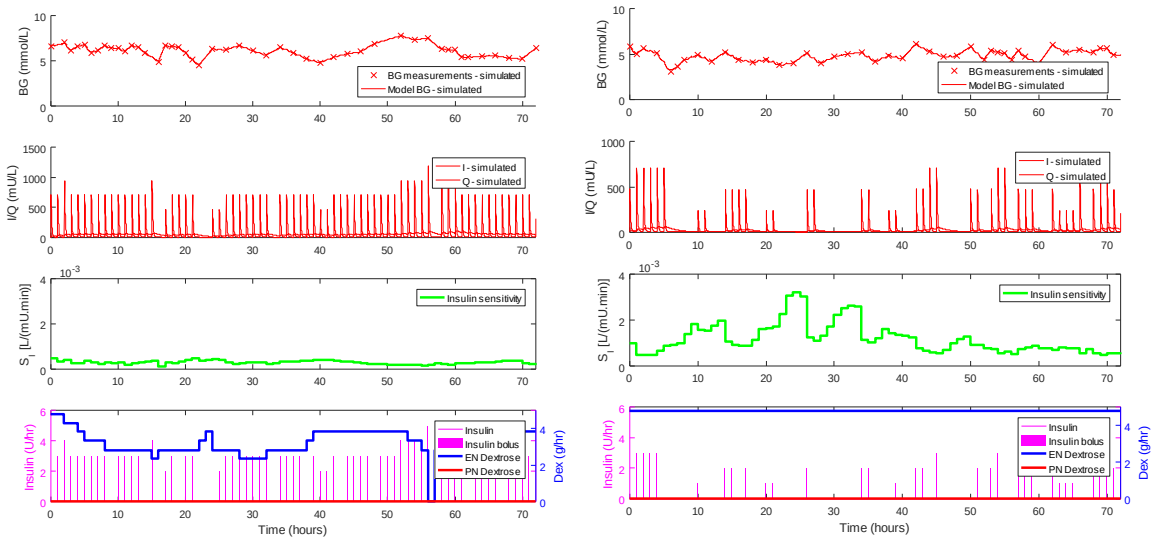

Figure A1-2: Example patients from clinical data, showing measured blood glucose (BG), clinically delivered insulin and nutrition, and model fitted insulin sensitivity (SI). A more 'Stable' SI profile (left) and more variable SI profile (right).

Appendix II: Clinical Significance Calculations

The equivalence range is determined by examining the changes in insulin sensitivity (SI) which become clinically significant. This is different from hypothesis testing, as the equivalence range is determined independently. Further, in theory, two samples can be clinically equivalent, even if they are statistically different. The equivalence range for insulin sensitivity changes can be determined in two ways:

- Change in SI possible due to measurement error
- Change in SI required to change the insulin dose recommendation

The first accounts for variation in model-based SI during the identification process due to measurement error, while the second examines its impact on GC outputs. The equivalence range will be determined conservatively as the minimum changes in SI required to cause clinically significant change, to ensure the strongest test of equivalence.

A2.1 Change in SI due to measurement error

Model-based blood glucose is defined:

$$\dot{G}(t) = -p_G G(t) - S_I(t)G(t)\frac{Q(t)}{1+\alpha_G Q(t)} + \frac{P(t) + EGP - CNS}{V_G} \quad \text{(A2.1)}$$

Parameter descriptions and values can be found in Table A2.1. Considering the average response over an hour, Equation 1 can be rewritten:

$$\frac{\Delta G}{60min} = -p_G G_m - S_I G_m \frac{Q_m}{1+\alpha_G Q_m} + \frac{P + EGP - CNS}{V_G} \quad \text{(A2.2)}$$

If Equation 2 is then re-arranged for SI:

$$S_I = \frac{(-p_G G_m + (P + EGP - CNS)/V_G - \Delta G/60min)(1+\alpha_G Q_m)}{G_m Q_m} \quad \text{(A2.3)}$$

For an absolute error in BG of ξ %, according to Equation 3 the new SI becomes

$$S_{I,\xi} = \frac{(-p_G G_m(1+\xi/100) + (P + EGP - CNS)/V_G - \Delta G/60min)(1+\alpha_G Q_m)}{Q_m G_m(1+\xi/100)} \quad \text{(A2.4)}$$

The possible percentage difference in SI due to measurement error is then:

$$\Delta S_{I,\xi} = \frac{S_{I,\xi} - S_I}{S_I} \quad \text{(A2.5)}$$

For the Arkray Glucocard X, a very similar device compared to the Super Glucocard II, the standard deviation of percentage error on a BG measurement is 9.4% [186]. From Equation 4, the insulin sensitivity can be affected by insulin and nutrition inputs (indirectly reflected through Q_m and P respectively), as well as the range of change of BG. The range of possible $\Delta S_{I,\xi}$ for different G_m values, Q_m values, $\Delta G/60min$, and nutrition inputs (P) was assessed.

Table A2.1 Glucose Parameter Values and Description.

Parameter	Value and/or Units	Description
G(t)	mmol/l	Blood glucose concentration
Q(t)	mU/l	Interstitial insulin concentration
S_I(t)	l/mU/min	Insulin sensitivity
$-p_G$	0.006/min	Kidney and general non-insulin mediated clearance
α_G	1/65 l/mU	Saturation of insulin-mediated glucose uptake
V_G	13.3 L	Glucose distribution volume
EGP	1.16 mmol/min	Endogenous glucose production (hepatic)
CNS	0.3 mmol/min	Glucose uptake by central nervous system

Nutrition range (P):

Assuming a goal feed of 2000 kCal/day (25 kCal/kg/day for an 80kg individual) of Glucerna (1.0 kCal/mL, 0.0812 g/mL glucose), which is the primary feed type in this unit and study, goal feed (rounded to the nearest 5mL/hr) is 85 mL/hr (38.3 mmol/hr glucose). The range of nutrition tested was 50:10:120% of this feed target.

Sensitivity to nutrition inputs is shown in Figure A2.1. As the nutrition increases, the percent change in SI required to account for glucometer error decreases due to higher overall insulin sensitivity.

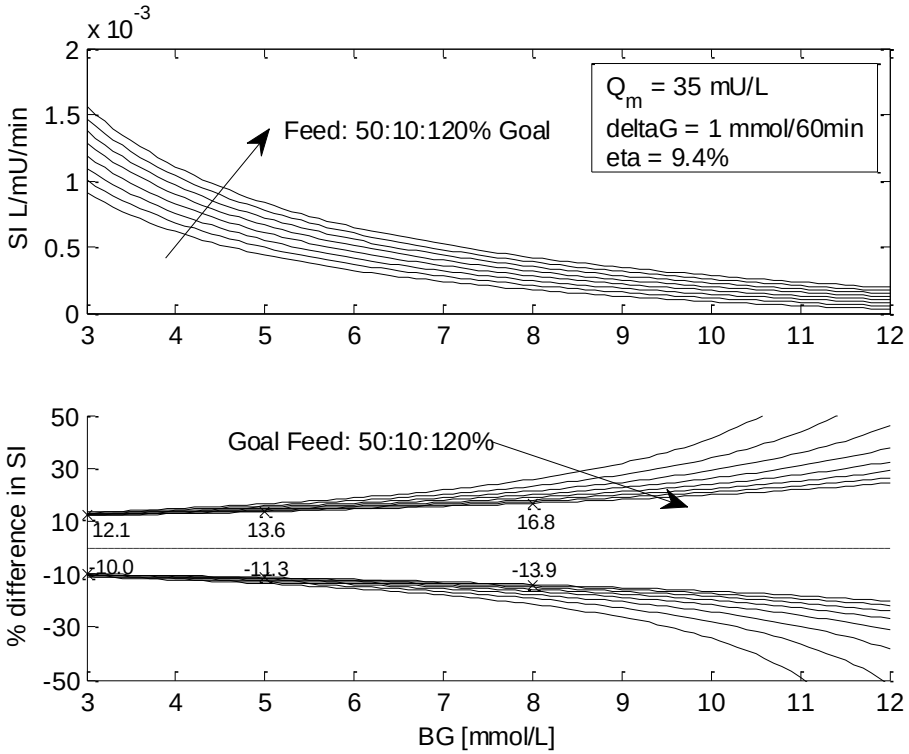

Figure A2.1: *Effect of nutrition inputs on insulin sensitivity (SI) and the potential difference in SI due to BG measurement error. Nutrition was modulated between 50 and 120% of the clinical goal (2000 kCal/day).*

Plasma insulin range (Q_m):

Sensitivity to Q_m within the common range (10 – 100 mU/L) was tested, and results are shown in Figure A2.2. The percentage difference in SI is not sensitive to Q_m, as it normalises in Equation 5. The case where a change in Q_m is causing the change in SI, rather than glucometer error as in Equations 4 and 5, is examined later in section A2.3.

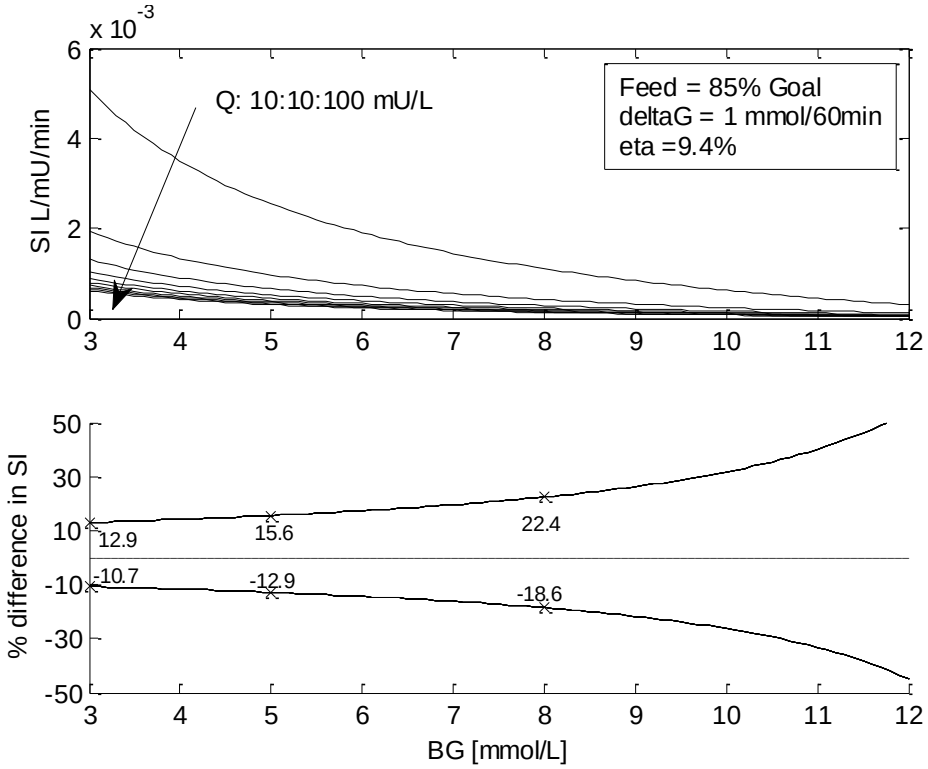

Figure A2.2: Effect of nutrition inputs on insulin sensitivity (SI) and the potential difference in SI due to BG measurement error. Interstitial insulin (Q) was modulated between 5 – 50 mU/L, which is a commonly observed range.

Rate of Change of Blood Glu*cose (ΔBG/60min)*:

Sensitivity to hourly changes in BG were examined, with this change ranging from an absolute change in BG of 2 mmol/L (a relatively large hourly change) to no change in BG. Results in Figure A2.3 show that the minimal change in SI required to account for glucometer error occurs at steady state (no change in BG). It is thus the most conservative case.

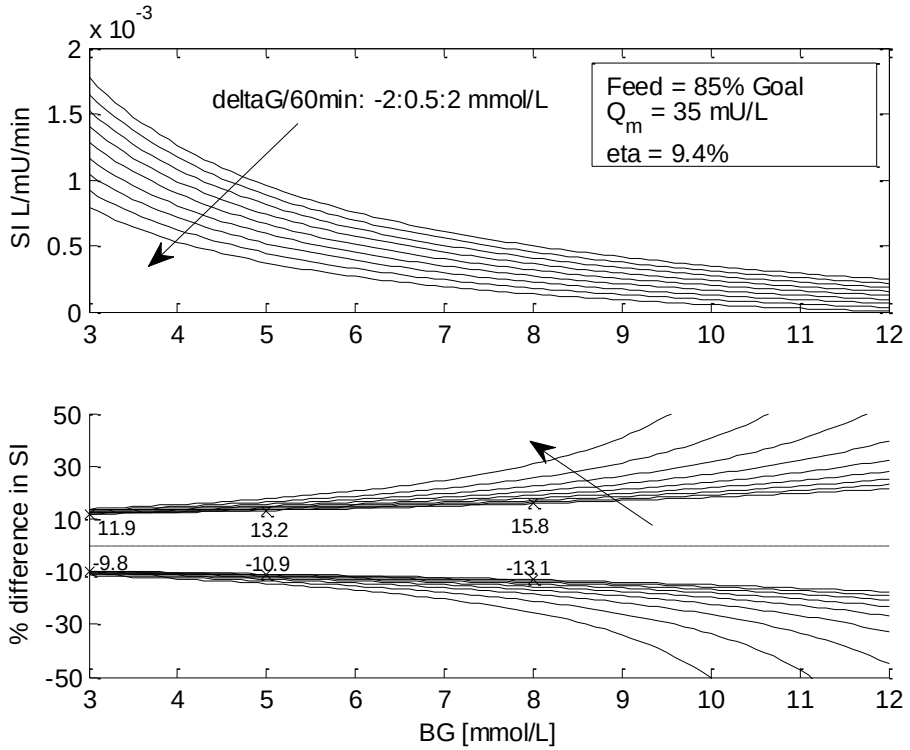

Figure A2.3: Effect of nutrition inputs on insulin sensitivity (SI) and the potential difference in SI due to BG measurement error. The change in BG over an hour was modulated between an absolute change of 2 mmol/L (a relatively large change) and no change.

The percentage change in SI required to account for glucometer error (ξ in %) is shown for a range of different error magnitudes in Figure A2.4. As the percentage or CV of glucometer error increases, the larger the clinically significant change in SI.

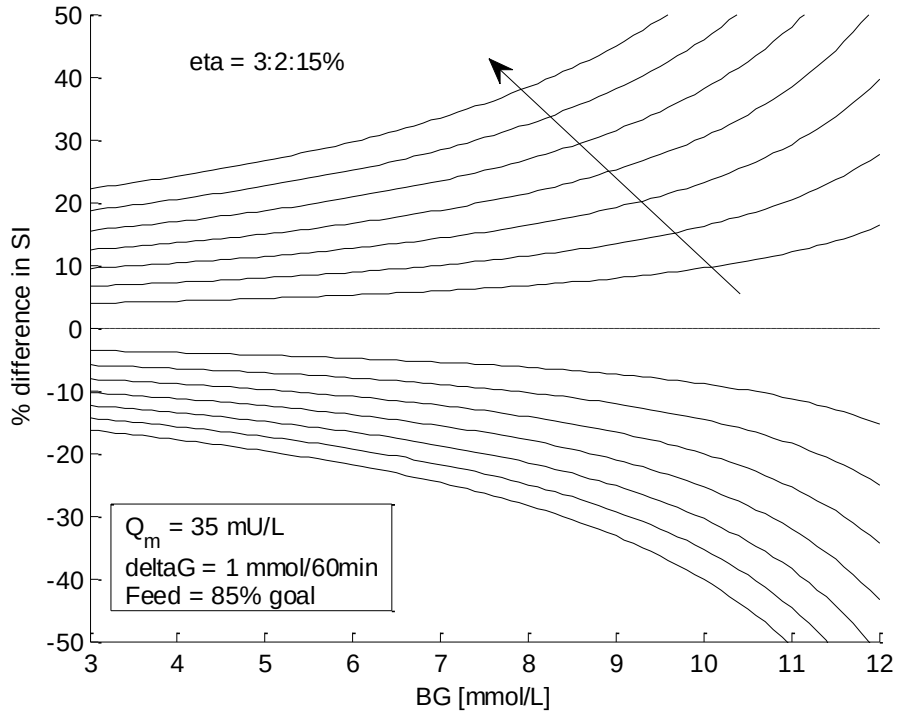

Figure A2.4: Effect of the magnitude of measurement error on the potential difference in SI due to BG measurement error.

A2.2 Change in SI required to change an insulin intervention

The plasma and interstitial insulin model equations are defined:

$$\dot{I}(t) = -n_K I(t) - n_L \frac{I(t)}{1 + \alpha_I I(t)} - n_I(I(t) - Q(t)) + \frac{u_{ex}(t)}{V_I} + (1 - x_L)\frac{u_{en}(G)}{V_I} \quad (6)$$

$$\dot{Q}(t) = n_I(I(t) - Q(t)) - n_C \frac{Q(t)}{1 + \alpha_G Q(t)} \quad (7)$$

Parameter descriptions and values are given in Table A2.2.

Table A2.2 Insulin Model Parameter Values and Description.

Parameter	Value and/or Units	Description
I(t)	mU/l	Plasma insulin concentration
Q(t)	mU/l	Interstitial insulin concentration
x_L	0.67	Fractional first pass hepatics insulin clearance from portal vein
n_L	0.1578 min^{-1}	Rate parameter: general hepatic insulin clearance
α_I	1.7x10-3 l/mU	Saturation of hepatics insulin clearance
n_K	0.0542 min^{-1}	Rate parameter: kidney clearance of insulin
n_C	0.006 min^{-1}	Rate parameter: cellular degradation of internalised insulin
n_I	0.006 min^{-1}	Rate parameter: diffusion of insulin between plasma and interstitium
u_{en}	mU/min	Pancreatic insulin secretion
V_I	4.0 L	Insulin distribution volume

Assuming steady state:

$$0 = -n_K I_{ss} - n_L \frac{I_{ss}}{1 + \alpha_I I_{ss}} - n_I(I_{ss} - Q_{ss}) + \frac{u_{ex}(t)}{V_I} + (1 - x_L)\frac{u_{en}(G)}{V_I} \quad (8)$$

$$0 = n_I(I_{ss} - Q_{ss}) - n_C \frac{Q_{ss}}{1 + \alpha_G Q_{ss}} \quad (9)$$

Equations 8 and 9 can be numerically solved (Newton's method was used here) for Qss and Iss for a given insulin dose, u_{ex}. As insulin secretion, u_{en}, is dependent on BG, this steady state will depend on the current BG level.

The change in insulin sensitivity required to change an insulin dose by 1 U, given a constant BG, can be estimated from the insulin-mediated glucose uptake component of Equation 1. The minimal degree

of allowable difference in SI for a clinical change in insulin treatments can be estimated from insulin-mediated glucose uptake and the assumption of no change in BG:

$$S_{I,1} G \frac{Q_{SS,1}}{1 + \alpha_G Q_{SS,1}} = S_{I,2} G \frac{Q_{SS,2}}{1 + \alpha_G Q_{SS,2}} \tag{10}$$

$$\frac{S_{I,2}}{S_{I,1}} = \frac{Q_{SS,1}}{Q_{SS,2}} \frac{\left(1 + \alpha_G Q_{SS,2}\right)}{\left(1 + \alpha_G Q_{SS,1}\right)} \tag{11}$$

The estimation of the minimal percentage change in SI required for a change in insulin intervention is shown in Figure A2.5. At higher insulin doses, Qss is higher and a 1 U change in insulin dose represents a smaller fraction of Qss. As a result, the percentage change in SI that allows a change in intervention narrows. Thus, the control system is more robust to error or variability in SI at lower insulin doses, which are more typical once BG is lowered into the normal range.

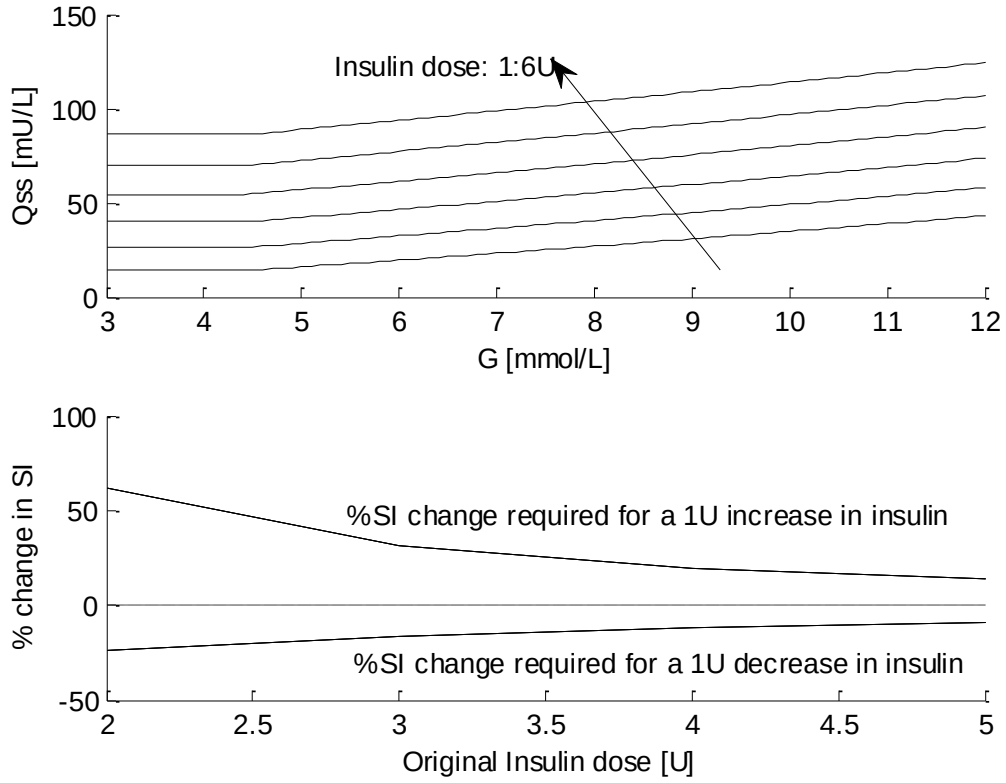

Figure A2.5: Model-based estimates of steady state interstitial insulin concentration for a given insulin dose, and the minimum percentage change in SI required to change an insulin intervention by 1U, assuming steady state Insulin.

A2.3 Summary and recommendation for equivalence

For a typical error standard deviation on Glucocard glucometers of 9.4%, to be conservative within the range of commonly observed BG (4.0 – 10.0 mmol/L) the minimum difference in SI that would be clinically significant beyond glucometer error is ~ 12-15% (Figures A2.1 – A2.3), and a function of BG. In the case of sustained hyperglycaemia, SI tends to be low, so percentage changes in SI required for clinical significance are higher.

Hence, equations 4 and 5 will be used to determine the equivalence range, using 100% goal feed and a conservative estimate of no change in blood glucose. The resulting plot is in Figure A2.6. For equivalence the 90% confidence interval (CI) must lie within the defined equivalence interval.

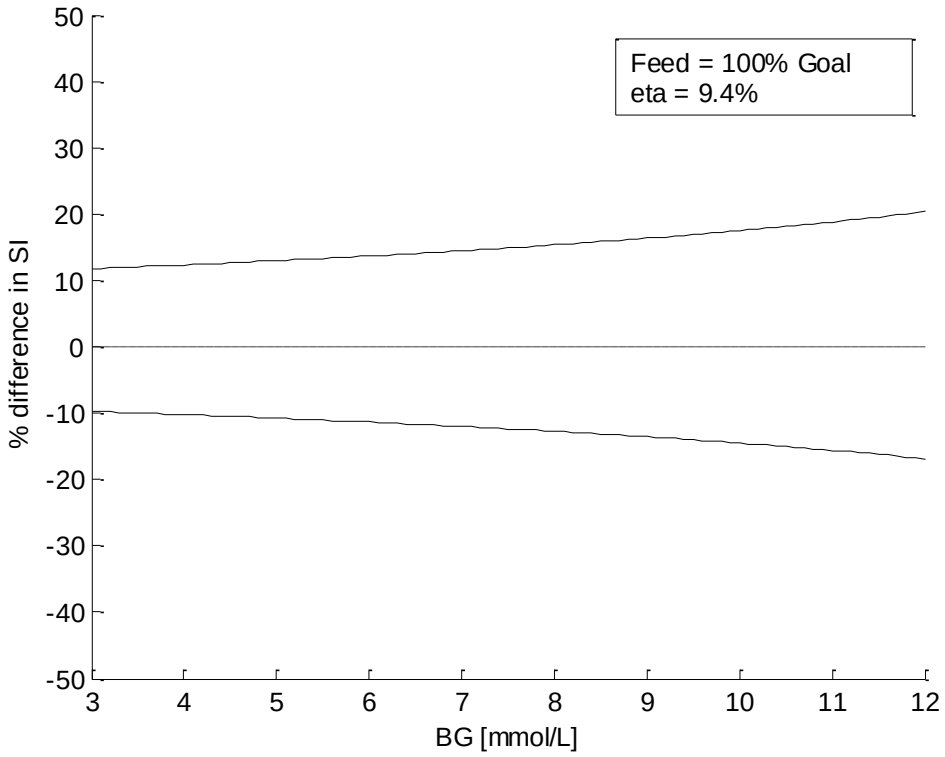

Figure A2.6: Equivalence range for insulin sensitivity (SI), using a conservative estimation of no change in blood glucose level (steady state). For non-steady state conditions, this range widens.

References

1. Guyton AC, Hall JE: Textbook of medical physiology. 10 edn. Philadelphie: London: Saunders; 2000.

2. Baron AD, Brechtel G, Wallace P, Edelman SV: **Rates and tissue sites of non-insulin- and insulin-mediated glucose uptake in humans.** Am J Physiol 1988, **255**(6 Pt 1):E769-774.

3. Mizock BA: **Alterations in fuel metabolism in critical illness: hyperglycaemia.** Best Pract Res Clin Endocrinol Metab 2001, **15**(4):533-551.

4. Weissman C: **The metabolic response to stress: an overview and update.** Anesthesiology 1990, **73**(2):308-327.

5. McCowen KC, Malhotra A, Bistrian BR: **Stress-induced hyperglycemia.** Crit Care Clin 2001, **17**(1):107-124.

6. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, et al: **Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012.** Intensive Care Med 2013, **39**(2):165-228.

7. Capes SE, Hunt D, Malmberg K, Gerstein HC: **Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview.** Lancet 2000, **355**(9206):773-778.

8. Kovalaske MA, Gandhi GY: **Glycemic control in the medical intensive care unit.** J Diabetes Sci Technol 2009, **3**(6):1330-1341.

9. Krinsley JS: **Association between hyperglycemia and increased hospital mortality in a heterogeneous population of critically ill patients.** Mayo Clin Proc 2003, **78**(12):1471-1478.

10. Bagshaw SM, Bellomo R, Jacka MJ, Egi M, Hart GK, George C: **The impact of early hypoglycemia and blood glucose variability on outcome in critical illness.** Crit Care 2009, **13**.

11. Egi M, Bellomo R, Stachowski E, French CJ, Hart GK, Taori G, Hegarty C, Bailey M: **Hypoglycemia and outcome in critically ill patients.** Mayo Clin Proc 2010, **85**(3):217-224.

12. Bistrian BR: **Hyperglycemia and infection: which is the chicken and which is the egg?** JPEN J Parenter Enteral Nutr 2001, **25**(4):180-181.

13. Dungan KM, Braithwaite SS, Preiser JC: **Stress hyperglycaemia.** Lancet 2009, **373**(9677):1798-1807.

14. Kalfon P, Le Manach Y, Ichai C, Brechot N, Cinotti R, Dequin PF, Riu-Poulenc B, Montravers P, Annane D, Dupont H, et al: **Severe and multiple hypoglycemic episodes are associated with increased risk of death in ICU patients.** Crit Care 2015, **19**:153.

15. American Diabetes Association: **Diagnosis and Classification of Diabetes Mellitus.** Diabetes Care 2013, **36(S1)**:S67-S74.

16. Diabetes Control Complications Trial Research Group, Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, Davis M, Rand L, Siebert C: **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus.** N Engl J Med 1993, **329**(14):977-986.

17. Wang PH, Lau J, Chalmers TC: **Meta-analysis of effects of intensive blood-glucose control on late complications of type I diabetes.** Lancet 1993, **341**(8856):1306-1309.

18. Reichard P, Nilsson BY, Rosenqvist U: **The effect of long-term intensified insulin treatment on the development of microvascular complications of diabetes mellitus.** N Engl J Med 1993, **329**(5):304-309.

19. American Diabetes Association: **Implications of the diabetes control and complications trial.** Diabetes Care 2003, **26 Suppl 1**:S25-27.

20. Stettler C, Allemann S, Juni P, Cull CA, Holman RR, Egger M, Krahenbuhl S, Diem P: **Glycemic control and macrovascular disease in types 1 and 2 diabetes mellitus: Meta-analysis of randomized trials.** Am Heart J 2006, **152**(1):27-38.

21. McCowen K, L.; K, B.R B: **Hyperglycemia and Blood Sugar Management: Implications for Infection.** In Nutrition Support to Pharmacologic Nutrition in the ICU Update in Intensive Care Medicine. Volume 34: Springer, Berlin, Heidelberg; 2002.

22. Marik PE, Raghavan M: **Stress-hyperglycemia, insulin and immunomodulation in sepsis.** Intensive Care Med 2004, **30**(5):748-756.

23. Preiser JC, Ichai C, Orban JC, Groeneveld AB: **Metabolic response to the stress of critical illness.** Br J Anaesth 2014, **113**(6):945-954.

24. Marik PE, Bellomo R: **Stress hyperglycemia: an essential survival response!** Crit Care 2013, **17**(2):305.

25. Luna B, Feinglos MN: **Drug-induced hyperglycemia.** JAMA 2001, **286**(16):1945-1948.

26. Pretty C, Chase JG, Lin J, Shaw GM, Le Compte A, Razak N, Parente JD: **Impact of glucocorticoids on insulin resistance in the critically ill.** Comput Methods Programs Biomed 2011, **102**(2):172-180.

27. Preiser JC, Fraipont V, Lheureux O: **The "baby stomach" concept applied to the nutrition of the critically ill.** Nutr Clin Metab 2019, **33**.

28. Arabi YM, Aldawood AS, Haddad SH, Al-Dorzi HM, Tamim HM, Jones G, Mehta S, McIntyre L, Solaiman O, Sakkijha MH, et al: **Permissive Underfeeding or Standard Enteral Feeding in Critically Ill Adults.** N Engl J Med 2015, **372**(25):2398-2408.

29. Arabi YM, Reintam Blaser A, Preiser JC: **Less is more in nutrition: critically ill patients are starving but not hungry.** Intensive Care Med 2019, **45**(11):1629-1631.

30. Weekers F, Giulietti AP, Michalaki M, Coopmans W, Van Herck E, Mathieu C, Van den Berghe G: **Metabolic, endocrine, and immune effects of stress hyperglycemia in a rabbit model of prolonged critical illness.** Endocrinology 2003, **144**(12):5329-5338.

31. Cely CM, Arora P, Quartin AA, Kett DH, Schein RM: **Relationship of baseline glucose homeostasis to hyperglycemia during medical critical illness.** Chest 2004, **126**(3):879-887.

32. Finfer S, Liu B, Chittock DR, Norton R, Myburgh JA, McArthur C, Mitchell I, Foster D, Dhingra V, Henderson WR, et al: **Hypoglycemia and risk of death in critically ill patients.** N Engl J Med 2012, **367**(12):1108-1118.

33. Krinsley JS, Grover A: **Severe hypoglycemia in critically ill patients: risk factors and outcomes.** Crit Care Med 2007, **35**(10):2262-2267.

34. Preiser JC, Brunkhorst F: **Tight glucose control and hypoglycemia.** Crit Care Med 2008, **36**(4):1391; author reply 1391-1392.

35. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R: **Intensive insulin therapy in critically ill patients.** N Engl J Med 2001, **345**(19):1359-1367.

36. Krinsley JS: **Effect of an intensive glucose management protocol on the mortality of critically ill adult patients.** Mayo Clin Proc 2004, **79**(8):992-1000.

37. Reed CC, Stewart RM, Sherman M, Myers JG, Corneille MG, Larson N, Gerhardt S, Beadle R, Gamboa C, Dent D, et al: **Intensive insulin protocol improves glucose control and is associated with a reduction in intensive care unit mortality.** J Am Coll Surg 2007, **204**(5):1048-1054; discussion 1054-1045.

38. Chase JG, Pretty CG, Pfeifer L, Shaw GM, Preiser JC, Le Compte AJ, Lin J, Hewett D, Moorhead KT, Desaive T: **Organ failure and tight glycemic control in the SPRINT study.** Crit Care 2010, **14**(4):R154.

39. Grey NJ, Perdrizet GA: **Reduction of nosocomial infections in the surgical intensive-care unit by strict glycemic control.** Endocr Pract 2004, **10 Suppl 2**:46-52.

40. Krinsley JS, Jones RL: **Cost analysis of intensive glycemic control in critically ill adult patients.** Chest 2006, **129**(3):644-650.

41. Blaha J, Kopecky P, Matias M, Hovorka R, Kunstyr J, Kotulak T, Lips M, Rubes D, Stritesky M, Lindner J, et al: **Comparison of three protocols for tight glycemic control in cardiac surgery patients.** Diabetes Care 2009, **32**(5):757-761.

42. Krinsley JS: **Is glycemic control of the critically ill cost-effective?** Hosp Pract (1995) 2014, **42**(4):53-58.

43. Mesotten D, Dubois J, Van Herpe T, van Hooijdonk RT, Wouters R, Coart D, Wouters P, Van Assche A, Veraghtert G, De Moor B, et al: **Software-guided versus nurse-directed blood glucose control in critically ill patients: the LOGIC-2 multicenter randomized controlled clinical trial.** Crit Care 2017, **21**(1):212.

44. Pachler C, Plank J, Weinhandl H, Chassin LJ, Wilinska ME, Kulnik R, Kaufmann P, Smolle KH, Pilger E, Pieber TR, et al: **Tight glycaemic control by an automated algorithm with time-variant sampling in medical ICU patients.** Intensive Care Med 2008, **34**(7):1224-1230.

45. Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, Milants I, Van Wijngaerden E, Bobbaers H, Bouillon R: **Intensive insulin therapy in the medical ICU.** N Engl J Med 2006, **354**(5):449-461.

46. Finfer S, Chittock DR, Su SY, Blair D, Foster D, Dhingra V, Bellomo R, Cook D, Dodek P, Henderson WR, et al: **Intensive versus conventional glucose control in critically ill patients.** N Engl J Med 2009, **360**(13):1283-1297.

47. Preiser JC, Devos P, Ruiz-Santana S, Melot C, Annane D, Groeneveld J, Iapichino G, Leverve X, Nitenberg G, Singer P, et al: **A prospective randomised multi-centre controlled trial on tight glucose control by intensive insulin therapy in adult intensive care units: the Glucontrol study.** Intensive Care Med 2009, **35**(10):1738-1748.

48. De La Rosa GC, Donado JH, Restrepo AH, Quintero AM, Gonzalez LG, Saldarriaga NE, Bedoya M, Toro JM, Velasquez JB, Valencia JC, et al: **Strict glycaemic control in patients hospitalised in a mixed medical and surgical intensive care unit: a randomised clinical trial.** Crit Care 2008, **12**(5):R120.

49. Arabi YM, Dabbagh OC, Tamim HM, Al-Shimemeri AA, Memish ZA, Haddad SH: **Intensive versus conventional insulin therapy: a randomized controlled trial in medical and surgical critically ill patients.** Crit Care Med 2008, **36**.

50. Brunkhorst FM, Engel C, Bloos F, Meier-Hellmann A, Ragaller M, Weiler N, Moerer O, Gruendling M, Oppert M, Grond S, et al: **Intensive insulin therapy and pentastarch resuscitation in severe sepsis.** N Engl J Med 2008, **358**(2):125-139.

51. Fahy BG, Sheehy AM, Coursin DB: **Glucose control in the intensive care unit.** Crit Care Med 2009, **37**.

52. Treggiari MM, Karir V, Yanez ND, Weiss NS, Daniel S, Deem SA: **Intensive insulin therapy and mortality in critically ill patients.** Crit Care 2008, **12**.

53. Preiser JC, Devos P, Chiolero R: **Which factors influence glycemic control in the intensive care unit?** Curr Opin Clin Nutr Metab Care 2010, **13**(2):205-210.

54. Ali NA, O'Brien JM, Dungan K, Phillips G, Marsh CB, Lemeshow S, Connors AF, Preiser JC: **Glucose variability and mortality in patients with sepsis.** Crit Care Med 2008, **36**(8):2316-2321.

55. Egi M, Bellomo R, Stachowski E, French CJ, Hart G: **Variability of blood glucose concentration and short-term mortality in critically ill patients.** Anesthesiology 2006, **105**(2):244-252.

56. Krinsley JS: **Glycemic variability and mortality in critically ill patients: the impact of diabetes.** J Diabetes Sci Technol 2009, **3**(6):1292-1301.

57. Donati A, Damiani E, Domizi R, Botticelli L, Castagnani R, Gabbanelli V: **Glycaemic variability, infections and mortality in a medical-surgical intensive care unit.** Crit Care Resusc 2014, **16**.

58. Kauffmann RM, Hayes RM, Buske BD, Norris PR, Campion TR, Dortch M: **Increasing blood glucose variability heralds hypoglycemia in the critically ill.** J Surg Res 2011, **170**.

59. Waeschle RM, Moerer O, Hilgers R, Herrmann P, Neumann P, Quintel M: **The impact of the severity of sepsis on the risk of hypoglycaemia and glycaemic variability.** Crit Care 2008, **12**.

60. Krinsley JS: **Is It Time to Rethink Blood Glucose Targets in Critically Ill Patients?** Chest 2018, **154**(5):1004-1005.

61. Preiser JC, Straaten HM: **Glycemic control: please agree to disagree.** Intensive Care Med 2016, **42**(9):1482-1484.

62. Marik PE: **Tight glycemic control in acutely ill patients: low evidence of benefit, high evidence of harm!** Intensive Care Med 2016, **42**(9):1475-1477.

63. Preiser JC: **NICE-SUGAR: the end of a sweet dream?** Crit Care 2009, **13**(3):143.

64. Chase JG, Dickson J: **Traversing the valley of glycemic control despair.** Critical Care 2017.

65. Griesdale DE, de Souza RJ, van Dam RM, Heyland DK, Cook DJ, Malhotra A, Dhaliwal R, Henderson WR, Chittock DR, Finfer S, Talmor D: **Intensive insulin therapy and mortality among critically ill patients: a meta-analysis including NICE-SUGAR study data.** CMAJ 2009, **180**(8):821-827.

66. Wiener RS, Wiener DC, Larson RJ: **Benefits and risks of tight glucose control in critically ill adults: a meta-analysis.** JAMA 2008, **300**(8):933-944.

67. Marik PE, Preiser JC: **Toward understanding tight glycemic control in the ICU: a systematic review and metaanalysis.** Chest 2010, **137**.

68. Krinsley J, Preiser JC: **Intensive insulin therapy to control hyperglycemia in the critically ill: a look back at the evidence shapes the challenges ahead.** Crit Care 2010, **14**(6):330.

69. Krinsley JS, Keegan MT: **Hypoglycemia in the critically ill: how low is too low?** Mayo Clin Proc 2010, **85**(3):215-216.

70. Singer P, Blaser AR, Berger MM, Alhazzani W, Calder PC, Casaer MP, Hiesmayr M, Mayer K, Montejo JC, Pichard C, et al: **ESPEN guideline on clinical nutrition in the intensive care unit.** Clin Nutr 2019, **38**(1):48-79.

71. Al-Tarifi A, Abou-Shala N, Tamim HM, Rishu AH, Arabi YM: **What is the optimal blood glucose target in critically ill patients? A nested cohort study.** Ann Thorac Med 2011, **6**.

72. Moghissi ES, Korytkowski MT, DiNardo M, Einhorn D, Hellman R, Hirsch IB, Inzucchi SE, Ismail-Beigi F, Kirkman MS, Umpierrez GE, et al: **American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control.** Endocr Pract 2009, **15**(4):353-369.

73. Jacobi J, Bircher N, Krinsley J, Agus M, Braithwaite SS, Deutschman C, Freire AX, Geehan D, Kohl B, Nasraway SA, et al: **Guidelines for the use of an insulin infusion for the management of hyperglycemia in critically ill patients.** Crit Care Med 2012, **40**(12):3251-3276.

74. Qaseem A, Humphrey LL, Chou R, Snow V, Shekelle P, Clinical Guidelines Committee of the American College of P: **Use of intensive insulin therapy for the management of glycemic control in hospitalized patients: a clinical practice guideline from the American College of Physicians.** Ann Intern Med 2011, **154**(4):260-267.

75. Vanhorebeek I, Gunst J, Van den Berghe G: **Critical Care Management of Stress-Induced Hyperglycemia.** Curr Diab Rep 2018, **18**(4):17.

76. Krinsley JS, Preiser JC: **Time in blood glucose range 70 to 140 mg/dl >80% is strongly associated with increased survival in non-diabetic critically ill adults.** Crit Care 2015, **19**:179.

77. Penning S, Chase JG, Preiser JC, Pretty CG, Signal M, Melot C, Desaive T: **Does the achievement of an intermediate glycemic target reduce organ failure and mortality? A post hoc analysis of the Glucontrol trial.** J Crit Care 2014, **29**(3):374-379.

78. Penning S, Pretty C, Preiser JC, Shaw GM, Desaive T, Chase JG: **Glucose control positively influences patient outcome: A retrospective study.** J Crit Care 2015, **30**(3):455-459.

79. Signal M, Le Compte A, Shaw GM, Chase JG: **Glycemic levels in critically ill patients: are normoglycemia and low variability associated with improved outcomes?** J Diabetes Sci Technol 2012, **6**(5):1030-1037.

80. Omar AS, Salama A, Allam M, Elgohary Y, Mohammed S, Tuli AK: **Association of time in blood glucose range with outcomes following cardiac surgery.** BMC anesthesiology 2015, **15**.

81. Chase JG, Shaw G, Le Compte A, Lonergan T, Willacy M, Wong XW, Lin J, Lotz T, Lee D, Hann C: **Implementation and evaluation of the SPRINT protocol for tight glycaemic control in critically ill patients: a clinical practice change.** Crit Care 2008, **12**(2):R49.

82. Pretty CG, Le Compte AJ, Chase JG, Shaw GM, Preiser JC, Penning S, Desaive T: **Variability of insulin sensitivity during the first 4 days of critical illness: implications for tight glycemic control.** Ann Intensive Care 2012, **2**(1):17.

83. Langouche L, Vander Perre S, Wouters PJ, D'Hoore A, Hansen TK, Van den Berghe G: **Effect of intensive insulin therapy on insulin sensitivity in the critically ill.** J Clin Endocrinol Metab 2007, **92**(10):3890-3897.

84. Pielmeier U, Rousing ML, Andreassen S, Nielsen BS, Haure P: **Decision support for optimized blood glucose control and nutrition in a neurotrauma intensive care unit: preliminary results of clinical advice and prediction accuracy of the Glucosafe system.** J Clin Monit Comput 2012, **26**.

85. Chase JG, Le Compte AJ, Suhaimi F, Shaw GM, Lynn A, Lin J, Pretty CG, Razak N, Parente JD, Hann CE, et al: **Tight glycemic control in critical care--the leading role of insulin sensitivity and patient variability: a review and model-based analysis.** Comput Methods Programs Biomed 2011, **102**(2):156-171.

86. Suhaimi F, Le Compte A, Preiser JC, Shaw GM, Massion P, Radermecker R, Pretty CG, Lin J, Desaive T, Chase JG: **What makes tight glycemic control tight? The impact of variability and nutrition in two clinical studies.** J Diabetes Sci Technol 2010, **4**(2):284-298.

87. Stewart KW, Pretty CG, Tomlinson H, Thomas FL, Homlok J, Noemi SN, Illyes A, Shaw GM, Benyo B, Chase JG: **Safety, efficacy and clinical generalization of the STAR protocol: a retrospective analysis.** Ann Intensive Care 2016, **6**(1):24.

88. Blaha J, Barteczko-Grajek B, Berezowicz P, Charvat J, Chvojka J, Grau T, Holmgren J, Jaschinski U, Kopecky P, Manak J, et al: **Space GlucoseControl system for blood glucose control in intensive care patients--a European multicentre observational study.** BMC anesthesiology 2016, **16**:8.

89. Kalfon P, Giraudeau B, Ichai C, Guerrini A, Brechot N, Cinotti R, Dequin PF, Riu-Poulenc B, Montravers P, Annane D, et al: **Tight computerized versus conventional glucose control in the ICU: a randomized controlled trial.** Intensive Care Med 2014, **40**(2):171-181.

90. Chase JG, Le Compte AJ, Preiser JC, Shaw GM, Penning S, Desaive T: **Physiological modeling, tight glycemic control, and the ICU clinician: what are models and how can they affect practice?** Ann Intensive Care 2011, **1**(1):11.

91. Chase JG, Preiser JC, Dickson JL, Pironet A, Chiew YS, Pretty CG, Shaw GM, Benyo B, Moeller K, Safaei S, et al: **Next-generation, personalised, model-based critical care medicine: a state-of-the art review of in silico virtual patient models, methods, and cohorts, and how to validation them.** Biomed Eng Online 2018, **17**(1):24.

92. Wong XW, Singh-Levett I, Hollingsworth LJ, Shaw GM, Hann CE, Lotz T, Lin J, Wong OS, Chase JG: **A novel, model-based insulin and nutrition delivery controller for glycemic regulation in critically ill patients.** Diabetes Technol Ther 2006, **8**(2):174-190.

93. Lin J, Razak NN, Pretty CG, Le Compte A, Docherty P, Parente JD, Shaw GM, Hann CE, Geoffrey Chase J: **A physiological Intensive Control Insulin-Nutrition-Glucose (ICING) model validated in critically ill patients.** Comput Methods Programs Biomed 2011, **102**(2):192-205.

94. Chase JG, Suhaimi F, Penning S, Preiser JC, Le Compte AJ, Lin J, Pretty CG, Shaw GM, Moorhead KT, Desaive T: **Validation of a model-based virtual trials method for tight glycemic control in intensive care.** Biomed Eng Online 2010, **9**:84.

95. Evans A, Le Compte A, Tan CS, Ward L, Steel J, Pretty CG, Penning S, Suhaimi F, Shaw GM, Desaive T, Chase JG: **Stochastic targeted (STAR) glycemic control: design, safety, and performance.** J Diabetes Sci Technol 2012, **6**(1):102-115.

96. Fisk LM, Le Compte AJ, Shaw GM, Penning S, Desaive T, Chase JG: **STAR development and protocol comparison.** IEEE Trans Biomed Eng 2012, **59**(12):3357-3364.

97. Chase JG, Shaw GM, Lin J, Doran CV, Hann C, Lotz T, Wake GC, Broughton B: **Targeted glycemic reduction in critical care using closed-loop control.** Diabetes Technol Ther 2005, **7**(2):274-282.

98. Chase JG, Benyo B, Desaive T: **Glycemic control in the intensive care unit: A control systems perspective.** Annual Reviews in Control 2019, **48**:359-368.

99. Chase JG, Desaive T, Bohe J, Cnop M, De Block C, Gunst J, Hovorka R, Kalfon P, Krinsley J, Renard E, Preiser JC: **Improving glycemic control in critically ill patients: personalized care to mimic the endocrine pancreas.** Crit Care 2018, **22**(1):182.

100. Pacini G, Bergman RN: **MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test.** Comput Methods Programs Biomed 1986, **23**(2):113-122.

101. Bergman RN, Ider YZ, Bowden CR, Cobelli C: **Quantitative estimation of insulin sensitivity.** Am J Physiol 1979, **236**(6):E667-677.

102. Mari A, Pacini G, Murphy E, Ludvik B, Nolan JJ: **A model-based method for assessing insulin sensitivity from the oral glucose tolerance test.** Diabetes Care 2001, **24**(3):539-548.

103. Blakemore A, Wang SH, Le Compte A, G MS, Wong XW, Lin J, Lotz T, C EH, Chase JG: **Model-based insulin sensitivity as a sepsis diagnostic in critical care.** J Diabetes Sci Technol 2008, **2**(3):468-477.

104. Le Compte AJ, Pretty CG, Lin J, Shaw GM, Lynn A, Chase JG: **Impact of variation in patient response on model-based control of glycaemia in critically ill patients.** Comput Methods Programs Biomed 2013, **109**(2):211-219.

105. Chase JG, Shaw GM, Lotz T, LeCompte A, Wong J, Lin J, Lonergan T, Willacy M, Hann CE: **Model-based insulin and nutrition administration for tight glycaemic control in critical care.** Curr Drug Deliv 2007, **4**(4):283-296.

106. Lonergan T, Compte AL, Willacy M, Chase JG, Shaw GM, Hann CE, Lotz T, Lin J, Wong XW: **A pilot study of the SPRINT protocol for tight glycemic control in critically Ill patients.** Diabetes Technol Ther 2006, **8**(4):449-462.

107. Hovorka R, Kremen J, Blaha J, Matias M, Anderlova K, Bosanska L, Roubicek T, Wilinska ME, Chassin LJ, Svacina S, Haluzik M: **Blood glucose control by a model predictive control algorithm with variable sampling rate versus a routine glucose management protocol in cardiac surgery patients: a randomized controlled trial.** J Clin Endocrinol Metab 2007, **92**(8):2960-2964.

108. Van Herpe T, Mesotten D, Wouters PJ, Herbots J, Voets E, Buyens J, De Moor B, Van den Berghe G: **LOGIC-insulin algorithm-guided versus nurse-directed blood glucose control during critical illness: the LOGIC-1 single-center, randomized, controlled clinical trial.** Diabetes Care 2013, **36**(2):188-194.

109. Amrein K, Ellmerer M, Hovorka R, Kachel N, Parcz D, Korsatko S, Smolle K, Perl S, Bock G, Doll W, et al: **Hospital glucose control: safe and reliable glycemic control using enhanced model predictive control algorithm in medical intensive care unit patients.** Diabetes Technol Ther 2010, **12**(5):405-412.

110. Motulsky H: Intuitive biostatistics: a nonmathematical guide to statistical thinking. New York: Oxford University Press; 2014.

111. Motulsky H: **Common misconceptions about data analysis and statistics.** Br J Pharmacol 2015, **172**.

112. Goodman SN: **Toward evidence-based medical statistics. 1: The P value fallacy.** Ann Intern Med 1999, **130**(12):995-1004.

113. Dixon P: **The p-value fallacy and how to avoid it.** Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale 2003, **57**(3):189-202.

114. Efron B, Tibshirani R: An Introduction to the Bootstrap. USA: Chapman & Hall / CRC; 1998.

115. Fethney J: **Statistical and clinical significance, and how to use confidence intervals to help interpret both.** Australian critical care : official journal of the Confederation of Australian Critical Care Nurses 2010, **23**(2):93-97.

116. Badawi O, Yeung SY, Rosenfeld BA: **Evaluation of glycemic control metrics for intensive care unit populations.** Am J Med Qual 2009, **24**(4):310-320.

117. Eslami S, de Keizer NF, de Jonge E, Schultz MJ, Abu-Hanna A: **A systematic review on quality indicators for tight glycaemic control in critically ill patients: need for an unambiguous indicator reference subset.** Crit Care 2008, **12**(6):R139.

118. Finfer S, Wernerman J, Preiser JC, Cass T, Desaive T, Hovorka R, Joseph JI, Kosiborod M, Krinsley J, Mackenzie I, et al: **Clinical review: Consensus recommendations on measurement of blood glucose and reporting glycemic control in critically ill adults.** Crit Care 2013, **17**(3):229.

119. Chase JG, Shaw GM, Hann CE, LeCompte A, Lonergan T, Willacy M, Wong XW, Lin J, Lotz T: **Clinical validation of a model-based glycaemic control design approach and comparison to other clinical protocols.** Conf Proc IEEE Eng Med Biol Soc 2006, **1**:59-62.

120. Chase JG, Shaw GM, Wong XW, Lotz T, Lin J, Hann CE: **Model-based glycaemic control in critical care - a review of the state of the possible.** Biomedical Signal Processing and Control 2006, **1**(1):3-21.

121. Mesotten D, Van den Berghe G: **Clinical benefits of tight glycaemic control: focus on the intensive care unit.** Best Pract Res Clin Anaesthesiol 2009, **23**(4):421-429.

122. Stewart KW, Pretty CG, Shaw GM, Chase JG: **Interpretation of Retrospective BG Measurements.** J Diabetes Sci Technol 2018, **12**(5):967-975.

123. Evans A, Shaw GM, Le Compte A, Tan CS, Ward L, Steel J, Pretty CG, Pfeifer L, Penning S, Suhaimi F, et al: **Pilot proof of concept clinical trials of Stochastic Targeted (STAR) glycemic control.** Ann Intensive Care 2011, **1**:38.

124. Lin J, Lee D, Chase JG, Shaw GM, Hann CE, Lotz T, Wong J: **Stochastic modelling of insulin sensitivity variability in critical care.** Biomedical Signal Processing and Control 2006, **1**(2):229-242.

125. Lin J, Lee D, Chase JG, Shaw GM, Le Compte A, Lotz T, Wong J, Lonergan T, Hann CE: **Stochastic modelling of insulin sensitivity and adaptive glycemic control for critical care.** Comput Methods Programs Biomed 2008, **89**(2):141-152.

126. Uyttendaele V, Dickson J, Stewart K, Desaive T, Benyo B, Szabo-Nemedi N, Illyes A, Shaw G, Chase G: **A 3D insulin sensitivity prediction model enables more patient-specific prediction and model-based glycaemic control.** Biomed Signal Process Control 2018.

127. Uyttendaele V, Knopp JL, Davidson S, Desaive T, Benyo B, Shaw GM, Chase JG: **3D kernel-density stochastic model for more personalized glycaemic control: development and in-silico validation.** BioMedical Engineering OnLine 2019, **18**(1):102.

128. Dickson JL, Stewart KW, Pretty CG, Flechet M, Desaive T, Penning S, Lambermont BC, Benyo B, Shaw GM, Chase JG: **Generalisability of a Virtual Trials Method for Glycaemic Control in Intensive Care.** IEEE Transactions on Biomedical Engineering 2017, **65**(7).

129. Lonergan T, Le Compte A, Willacy M, Chase JG, Shaw GM, Wong XW, Lotz T, Lin J, Hann CE: **A simple insulin-nutrition protocol for tight glycemic control in critical illness: development and protocol comparison.** Diabetes Technol Ther 2006, **8**(2):191-206.

130. Docherty PD, Chase JG, David T: **Characterisation of the iterative integral parameter identification method.** Med Biol Eng Comput 2012, **50**(2):127-134.

131. Hann CE, Chase JG, Lin J, Lotz T, Doran CV, Shaw GM: **Integral-based parameter identification for long-term dynamic verification of a glucose-insulin system model.** Comput Methods Programs Biomed 2005, **77**(3):259-270.

132. Natali A, Gastaldelli A, Camastra S, Sironi AM, Toschi E, Masoni A, Ferrannini E, Mari A: **Dose-response characteristics of insulin action on glucose metabolism: a non-steady-state approach.** Am J Physiol Endocrinol Metab 2000, **278**(5):E794-801.

133. Prigeon RL, Roder ME, Porte D, Jr., Kahn SE: **The effect of insulin dose on the measurement of insulin sensitivity by the minimal model technique. Evidence for saturable insulin transport in humans.** J Clin Invest 1996, **97**(2):501-507.

134. Stewart KW, Chase JG, Pretty CG, Shaw GM: **Nutrition delivery of a model-based ICU glycaemic control system.** Ann Intensive Care 2018, **8**(1):4.

135. Uyttendaele V, Dickson JL, Shaw GM, Desaive T, Chase JG: **Untangling glycaemia and mortality in critical care.** Crit Care 2017, **21**(1):152.

136. Docherty PD, Chase JG, Lotz TF, Hann CE, Shaw GM, Berkeley JE: **Independent cohort cross-validation of the real-time DISTq estimation of insulin sensitivity.** Comput Methods Programs Biomed 2011, **102**.

137. McAuley KA, Berkeley JE, Docherty PD, Lotz TF, Te Morenga LA, Shaw GM, Williams SM, Chase JG, Mann JI: **The dynamic insulin sensitivity and secretion test - a novel measure of insulin sensitivity.** Metabolism 2011, **60**(12):1748-1756.

138. Lotz TF, Chase JG, McAuley KA, Lee DS, Lin J, Hann CE, Mann JI: **Transient and steady-state euglycemic clamp validation of a model for glycemic control and insulin sensitivity testing.** Diabetes Technol Ther 2006, **8**(3):338-346.

139. Othman NA, Docherty PD, Krebs JD, Bell DA, Chase JG: **The necessity of identifying the basal glucose set-point in the IVGTT for patients with Type 2 Diabetes.** Biomed Eng Online 2015, **14**:18.

140. Lotz TF, Chase JG, McAuley KA, Shaw GM, Wong XW, Lin J, Lecompte A, Hann CE, Mann JI: **Monte Carlo analysis of a new model-based method for insulin sensitivity testing.** Comput Methods Programs Biomed 2008, **89**(3):215-225.

141. Pretty CG, Le Compte A, Penning S, Fisk L, Shaw GM, Desaive T, Chase JG: **Interstitial insulin kinetic parameters for a 2-compartment insulin model with saturable clearance.** Comput Methods Programs Biomed 2014, **114**(3):e39-45.

142. Silverman BW: Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.

143. Krishnan JA, Parce PB, Martinez A, Diette GB, Brower RG: **Caloric intake in medical ICU patients: consistency of care with guidelines and relationship to clinical outcomes.** Chest 2003, **124**(1):297-305.

144. Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T: **Randomized controlled trials - a matter of design.** Neuropsychiatr Dis Treat 2016, **12**:1341-1349.

145. Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, Cooney E, Mikkelsen ME, Kerlin MP, Small DS, Halpern SD: **Outcomes and statistical power in adult critical care randomized trials.** Am J Respir Crit Care Med 2014, **189**(12):1469-1478.

146. Vincent JL: **Improved survival in critically ill patients: are large RCTs more useful than personalized medicine? No.** Intensive Care Med 2016, **42**(11):1778-1780.

147. Vincent JL: **We should abandon randomized controlled trials in the intensive care unit.** Crit Care Med 2010, **38**(10 Suppl):S534-538.

148. Fernandez A, Sturmberg J, Lukersmith S, Madden R, Torkfar G, Colagiuri R, Salvador-Carulla L: **Evidence-based medicine: is it a bridge too far?** Health Res Policy Syst 2015, **13**:66.

149. Penning S, Le Compte AJ, Massion P, Moorhead KT, Pretty CG, Preiser JC, Shaw GM, Suhaimi F, Desaive T, Chase JG: **Second pilot trials of the STAR-Liege protocol for tight glycemic control in critically ill patients.** Biomed Eng Online 2012, **11**:58.

150. Knopp JL, Lynn AM, Shaw GM, Chase JG: **Safe and effective glycaemic control in premature infants: observational clinical results from the computerised STAR-GRYPHON protocol.** Arch Dis Child Fetal Neonatal Ed 2019, **104**(2):F205-F211.

151. Penning S, Le Compte AJ, Moorhead KT, Desaive T, Massion P, Preiser JC, Shaw GM, Chase JG: **First pilot trial of the STAR-Liege protocol for tight glycemic control in critically ill patients.** Comput Methods Programs Biomed 2012, **108**(2):844-859.

152. Chase JG, Andreassen S, Jensen K, Shaw GM: **Impact of human factors on clinical protocol performance: a proposed assessment framework and case examples.** J Diabetes Sci Technol 2008, **2**(3):409-416.

153. Dickson J, LeCompte A, Floyd RP, Chase JG, Lynn A, Shaw G: **Development and optimisation of stochastic targeted (STAR) glycaemic control for pre-term infants in neonatal intensive care.** Biomedical Signal Processing and Control 2012.

154. Uyttendaele V, Knopp JL, Pirotte M, Morimont P, Lambermont B, Shaw GM, Desaive T, Chase JG: **STAR-Liège Clinical Trial Interim Results: Safe and Effective Glycemic Control for All.** In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin, Germany: IEEE; 2019: 277-280.

155. Docherty PD, Chase JG, Lotz T, Hann CE, Shaw GM, Berkeley JE: **DISTq: an iterative analysis of glucose data for low-cost, real-time and accurate estimation of insulin sensitivity.** Open Med Inform J 2009, **3**.

156. Lotz TF, Chase JG, McAuley KA, Shaw GM, Docherty PD, Berkeley JE, Williams SM, Hann CE, Mann JI: **Design and clinical pilot testing of the model-based dynamic insulin sensitivity and secretion test (DISST).** J Diabetes Sci Technol 2010, **4**(6):1408-1423.

157. Le Compte A, Chase JG, Lynn A, Hann C, Shaw G, Wong XW, Lin J: **Blood glucose controller for neonatal intensive care: virtual trials development and first clinical trials.** J Diabetes Sci Technol 2009, **3**(5):1066-1081.

158. Jamaludin UK, F MS, Abdul Razak NN, Md Ralib A, Mat Nor MB, Pretty CG, Humaidi L: **Performance of Stochastic Targeted Blood Glucose Control Protocol by virtual trials in the Malaysian intensive care unit.** Comput Methods Programs Biomed 2018, **162**:149-155.

159. Uyttendaele V, Dickson JL, Shaw GM, Desaive T, Chase JG: **Virtual Trials of the NICE-SUGAR Protocol: The Impact on Performance of Protocol and Protocol Compliance.** IFAC-PapersOnline 2017, **50**(1):6672-6677.

160. Uyttendaele V, Knopp JL, Shaw GM, Desaive T, Chase JG: **Is intensive insulin therapy the scapegoat for or cause of hypoglycaemia and poor outcome?** IFAC Journal of Systems and Control 2019, **9**.

161. Krinsley JS: **Glucose control reduces ICU stay and mortality.** Perform Improv Advis 2005, **9**(1):4-6, 1.

162. Chase JG, Hann CE, Shaw GM, Wong J, Lin J, Lotz T, Lecompte A, Lonergan T: **Overview of glycemic control in critical care: relating performance and clinical results.** J Diabetes Sci Technol 2007, **1**(1):82-91.

163. Krinsley JS, Schultz MJ, Spronk PE, Harmsen RE, Braam HF, Sluijs JP: **Mild hypoglycemia is independently associated with increased mortality in the critically ill.** Crit Care 2011, **15**.

164. Van den Berghe G, Schetz M, Vlasselaers D, Hermans G, Wilmer A, Bouillon R, Mesotten D: **Clinical review: Intensive insulin therapy in critically ill patients: NICE-SUGAR or Leuven blood glucose target?** J Clin Endocrinol Metab 2009, **94**(9):3163-3170.

165. Krinsley J: **Glycemic control in critically ill patients: Leuven and beyond.** Chest 2007, **132**(1):1-2.

166. Krinsley JS: **Glycemic control in the critically ill: What have we learned since NICE-SUGAR?** Hosp Pract (1995) 2015, **43**(3):191-197.

167. Krinsley JS, Preiser JC: **Moving beyond tight glucose control to safe effective glucose control.** Crit Care 2008, **12**(3):149.

168. Brunkhorst FM, Reinhart K: **Intensive insulin therapy in the ICU: benefit versus harm?** Intensive Care Med 2007, **33**(7):1302.

169. Krinsley JS, Bruns DE, Boyd JC: **The impact of measurement frequency on the domains of glycemic control in the critically ill--a Monte Carlo simulation.** J Diabetes Sci Technol 2015, **9**(2):237-245.

170. Pretty CG, Signal M, Fisk L, Penning S, Le Compte A, Shaw GM, Desaive T, Chase JG: **Impact of sensor and measurement timing errors on model-based insulin sensitivity.** Comput Methods Programs Biomed 2014, **114**(3):e79-86.

171. Hersh AM, Hirshberg EL, Wilson EL, Orme JF, Morris AH, Lanspa MJ: **Lower Glucose Target Is Associated With Improved 30-Day Mortality in Cardiac and Cardiothoracic Patients.** Chest 2018, **154**(5):1044-1051.

172. Schultz MJ, Harmsen RE, Korevaar JC, Abu-Hanna A, Van Braam Houckgeest F, Van Der Sluijs JP, Spronk PE: **Adoption and implementation of the original strict glycemic control guideline is feasible and safe in adult critically ill patients.** Minerva Anestesiol 2012, **78**(9):982-995.

173. **The NICE-SUGAR Study** [https://studies.thegeorgeinstitute.org/nice/]

174. Aragon D: **Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control.** Am J Crit Care 2006, **15**(4):370-377.

175. Holzinger U, Feldbacher M, Bachlechner A, Kitzberger R, Fuhrmann V, Madl C: **Improvement of glucose control in the intensive care unit: an interdisciplinary collaboration study.** Am J Crit Care 2008, **17**(2):150-156.

176. Anger KE, Szumita PM: **Barriers to glucose control in the intensive care unit.** Pharmacotherapy 2006, **26**(2):214-228.

177. Schultz MJ, de Graaff MJ, Royakkers AA, van Braam Houckgeest F, van der Sluijs JP, Kieft H, Spronk PE: **Practice of strict glycemic control in critically ill patients.** Med Sci Monit 2008, **14**(11):RA191-197.

178. Schultz MJ, Harmsen RE, Spronk PE: **Clinical review: Strict or loose glycemic control in critically ill patients--implementing best available evidence from randomized controlled trials.** Crit Care 2010, **14**(3):223.

179. Carayon P, Gurses AP: **A human factors engineering conceptual framework of nursing workload and patient safety in intensive care units.** Intensive Crit Care Nurs 2005, **21**(5):284-301.

180. Collier B, Dossett LA, May AK, Diaz JJ: **Glucose control and the inflammatory response.** Nutr Clin Pract 2008, **23**(1):3-15.

181. Van den Berghe G, Wouters PJ, Bouillon R, Weekers F, Verwaest C, Schetz M, Vlasselaers D, Ferdinande P, Lauwers P: **Outcome benefit of intensive insulin therapy in the critically ill: Insulin dose versus glycemic control.** Crit Care Med 2003, **31**(2):359-366.

182. Egi M, Finfer S, Bellomo R: **Glycemic control in the ICU.** Chest 2011, **140**.

183. Krinsley JS: **Glycemic variability: a strong independent predictor of mortality in critically ill patients.** Crit Care Med 2008, **36**.

184. Thomas F, Pretty CG, Fisk L, Shaw GM, Chase JG, Desaive T: **Reducing the impact of insulin sensitivity variability on glycaemic outcomes using separate stochastic models within the STAR glycaemic protocol.** Biomed Eng Online 2014, **13**:43.

185. Le Compte AJ, Lee DS, Chase JG, Lin J, Lynn A, Shaw GM: **Blood glucose prediction using stochastic modeling in neonatal intensive care.** IEEE Trans Biomed Eng 2010, **57**(3):509-518.

186. Freckmann G, Baumstark A, Jendrike N, Zschornack E, Kocher S, Tshiananga J, Heister F, Haug C: **System accuracy evaluation of 27 blood glucose monitoring systems according to DIN EN ISO 15197.** Diabetes Technol Ther 2010, **12**(3):221-231.

187. SKUP: **Glucocard X-Meter : Meter and test strips designed for glucose self-measurement manufactered by Arkray, Inc.** In. Norway: University of Bergen; 2006.

188. Jamaludin UK, Docherty PD, Chase JG, Compte A, Shaw GM, Desaive T: **Observation of incretin effects during enteral feed transitions of critically ill patients.** ESPEN J 2012, **7**.

189. Jamaludin UK, Docherty PD, Chase JG, Shaw GM: **Impact of Haemodialysis on Insulin Kinetics of Acute Kidney Injury Patients in Critical Care.** J Med Biol Eng 2015, **35**(1):125-133.

190. Sah Pri A, Chase JG, Pretty CG, Shaw GM, Preiser JC, Vincent JL, Oddo M, Taccone FS, Penning S, Desaive T: **Evolution of insulin sensitivity and its variability in out-of-hospital cardiac arrest (OHCA) patients treated with hypothermia.** Crit Care 2014, **18**(5):586.

191. Signal M, Fisk L, Shaw GM, Chase JG: **Concurrent Continuous Glucose Monitoring in Critically Ill Patients: Interim Results and Observations.** J Diabetes Sci Technol 2013, **7**(6):1652-1653.

192. Signal M, Pretty CG, Chase JG, Le Compte A, Shaw GM: **Continuous glucose monitors and the burden of tight glycemic control in critical care: can they cure the time cost?** J Diabetes Sci Technol 2010, **4**(3):625-635.

193. Facchinetti A, Del Favero S, Sparacino G, Castle JR, Ward WK, Cobelli C: **Modeling the glucose sensor error.** IEEE Trans Biomed Eng 2014, **61**(3):620-629.

194. Reifman J, Rajaraman S, Gribok A, Ward WK: **Predictive monitoring for improved management of glucose levels.** J Diabetes Sci Technol 2007, **1**(4):478-486.

195. Zimmermann JB, Lehmann M, Hofer S, Husing J, Alles C, Werner J, Stiller J, Kunnecke W, Luntz S, Motsch J, Weigand MA: **Design of a prospective clinical study on the agreement between the Continuous GlucoseMonitor, a novel device for CONTinuous ASSessment of blood GLUcose levels, and the RAPIDLab(R) 1265 blood gas analyser: The CONTASSGLU study.** BMC anesthesiology 2012, **12**:24.

196. Breton M, Kovatchev B: **Analysis, modeling, and simulation of the accuracy of continuous glucose sensors.** J Diabetes Sci Technol 2008, **2**(5):853-862.

197. Kuure-Kinsey M, Palerm CC, Bequette BW: **A dual-rate Kalman filter for continuous glucose monitoring.** Conf Proc IEEE Eng Med Biol Soc 2006, **1**:63-66.

198. Lunn DJ, Wei C, Hovorka R: **Fitting dynamic models with forcing functions: application to continuous glucose monitoring in insulin therapy.** Stat Med 2011, **30**(18):2234-2250.

199. Brunner R, Adelsmayr G, Herkner H, Madl C, Holzinger U: **Glycemic variability and glucose complexity in critically ill patients: a retrospective analysis of continuous glucose monitoring data.** Crit Care 2012, **16**(5):R175.

200. Lundelin K, Vigil L, Bua S, Gomez-Mestre I, Honrubia T, Varela M: **Differences in complexity of glycemic profile in survivors and nonsurvivors in an intensive care unit: a pilot study.** Crit Care Med 2010, **38**(3):849-854.

201. Thomas F, Signal M, Chase JG: **Using Continuous Glucose Monitoring Data and Detrended Fluctuation Analysis to Determine Patient Condition: A Review.** J Diabetes Sci Technol 2015, **9**(6):1327-1335.

202. Signal M, Thomas F, Shaw GM, Chase JG: **Complexity of continuous glucose monitoring data in critically ill patients: continuous glucose monitoring devices, sensor locations, and detrended fluctuation analysis methods.** J Diabetes Sci Technol 2013, **7**(6):1492-1506.

203. Kinasewitz GT, Yan SB, Basson B, Comp P, Russell JA, Cariou A, Um SL, Utterback B, Laterre PF, Dhainaut JF, Group PSS: **Universal changes in biomarkers of coagulation and inflammation occur in patients with severe sepsis, regardless of causative micro-organism [ISRCTN74215569].** Crit Care 2004, **8**(2):R82-90.

204. Pachaly MA, do Nascimento MM, Suliman ME, Hayashi SY, Riella MC, Manfro RC, Stenvinkel P, Lindholm B: **Interleukin-6 is a better predictor of mortality as compared to C-reactive protein, homocysteine, pentosidine and advanced oxidation protein products in hemodialysis patients.** Blood purification 2008, **26**(2):204-210.

205. Hall MW, Geyer SM, Guo CY, Panoskaltsis-Mortari A, Jouvet P, Ferdinands J, Shay DK, Nateri J, Greathouse K, Sullivan R, et al: **Innate immune function and mortality in critically ill children with influenza: a multicenter study.** Crit Care Med 2013, **41**(1):224-236.

206. Stanojcic M, Chen P, Xiu F, Jeschke MG: **Impaired Immune Response in Elderly Burn Patients: New Insights Into the Immune-senescence Phenotype.** Ann Surg 2016, **264**(1):195-202.

207. Cabrera-Cancio MR: **Infections and the compromised immune status in the chronically critically ill patient: prevention strategies.** Respiratory care 2012, **57**(6):979-990; discussion 990-972.

208. Kumar M, Roe K, Nerurkar PV, Namekar M, Orillo B, Verma S, Nerurkar VR: **Impaired virus clearance, compromised immune response and increased mortality in type 2 diabetic mice infected with West Nile virus.** PLoS One 2012, **7**(8):e44682.

209. Langouche L, Vanhorebeek I, Van den Berghe G: **The role of insulin therapy in critically ill patients.** Treat Endocrinol 2005, **4**(6):353-360.

210. Butler SO, Btaiche IF, Alaniz C: **Relationship between hyperglycemia and infection in critically ill patients.** Pharmacotherapy 2005, **25**(7):963-976.

211. Fernandez-Real JM, Broch M, Richart C, Vendrell J, Lopez-Bermejo A, Ricart W: **CD14 monocyte receptor, involved in the inflammatory cascade, and insulin sensitivity.** J Clin Endocrinol Metab 2003, **88**(4):1780-1784.

212. Spindler MP, Ho AM, Tridgell D, McCulloch-Olson M, Gersuk V, Ni C, Greenbaum C, Sanda S: **Acute hyperglycemia impairs IL-6 expression in humans.** Immun Inflamm Dis 2016, **4**(1):91-97.

213. Turina M, Fry DE, Polk HC: **Acute hyperglycemia and the innate immune system: clinical, cellular, and molecular aspects.** Crit Care Med 2005, **33**.

214. Krogh-Madsen R, Moller K, Dela F, Kronborg G, Jauffred S, Pedersen BK: **Effect of hyperglycemia and hyperinsulinemia on the response of IL-6, TNF-alpha, and FFAs to low-dose endotoxemia in humans.** Am J Physiol Endocrinol Metab 2004, **286**(5):E766-772.

215. Dandona P, Mohanty P, Chaudhuri A, Garg R, Aljada A: **Insulin infusion in acute illness.** J Clin Invest 2005, **115**(8):2069-2072.

216. Vanhorebeek I, Langouche L, Van den Berghe G: **Glycemic and nonglycemic effects of insulin: how do they contribute to a better outcome of critical illness?** Curr Opin Crit Care 2005, **11**(4):304-311.

217. Chase JG, Andreassen S, Pielmeier U, Hann CE, McAuley KA, Mann JI: **A glucose-insulin pharmacodynamic surface modeling validation and comparison of metabolic system models.** Biomed Signal Process Control 2009, **4**.

218. Dickson JL, Chase JG, Gunn CA, Pretty C, Lynn A, Alsweiler J: **Gender and glycaemia: Insulin sensitivity and secretion in premature neonates.** IFAC Proceedings Volumes 2014, **47**(3):10168-10173.

219. Dickson JL, Alsweiler J, Gunn CA, Pretty CG, Chase JG: **A C-Peptide-Based Model of Pancreatic Insulin Secretion in Extremely Preterm Neonates in Intensive Care.** J Diabetes Sci Technol 2015, **10**(1):111-118.

220. Dickson JL, Chase JG, Pretty CG, Gunn CA, Alsweiler JM: **Hyperglycaemic Preterm Babies Have Sex Differences in Insulin Secretion.** Neonatology 2015, **108**(2):93-98.

221. Uyttendaele V, Knopp JL, Gottlieb R, Shaw GM, Desaive T, Chase JG: **Insulin Resistance in ICU Patients: Women Have Stronger Metabolic Response.** IFAC-PapersOnline 2020, **2020**.

222. Kwiatkowski K, Coe K, Bailar JC, Swanson GM: **Inclusion of minorities and women in cancer clinical trials, a decade later: Have we improved?** Cancer 2013, **119**(16):2956-2963.

223. Schiebinger L: **Women's health and clinical trials.** J Clin Invest 2003, **112**(7):973-977.

224. Sherman LA, Temple R, Merkatz RB: **Women in clinical trials: an FDA perspective.** Science 1995, **269**(5225):793-795.

225. Ruiz Cantero MT, Angeles Pardo M: **European Medicines Agency policies for clinical trials leave women unprotected.** J Epidemiol Community Health 2006, **60**(11):911-913.

226. Office UGA: **Women's health: FDA needs to ensure more study of gender differences in prescription drug testing.** In; 1992: 39.[vol GAO/HRD-93-17].

227. Merkatz RB, Temple R, Subel S, Feiden K, Kessler DA: **Women in clinical trials of new drugs. A change in Food and Drug Administration policy. The Working Group on Women in Clinical Trials.** N Engl J Med 1993, **329**(4):292-296.

228. Jackson G: Pain and Prejudice. Australia: Allen and Unwin; 2019.

229. Cerra FB, Benitez MR, Blackburn GL, Irwin RS, Jeejeebhoy K, Katz DP, Pingleton SK, Pomposelli J, Rombeau JL, Shronts E, et al: **Applied nutrition in ICU patients. A consensus statement of the American College of Chest Physicians.** Chest 1997, **111**(3):769-778.

230. Lheureux O, Preiser JC: **Year in review 2013: Critical Care--metabolism.** Crit Care 2014, **18**(5):571.

231. Losser MR, Damoisel C, Payen D: **Bench-to-bedside review: Glucose and stress conditions in the intensive care unit.** Crit Care 2010, **14**(4):231.

232. Nechas E, Foley F: Unequal treatment : what you don't know about how women are mistreated by the medical community. New York, USA: Simon & Schuste; 1994.

233. Flanagan DE, Holt RI, Owens PC, Cockington RJ, Moore VM, Robinson JS, Godsland IF, Phillips DI: **Gender differences in the insulin-like growth factor axis response to a glucose load.** Acta physiologica 2006, **187**(3):371-378.

234. Geer EB, Shen W: **Gender differences in insulin resistance, body composition, and energy balance.** Gender medicine 2009, **6 Suppl 1**:60-75.

235. Soeters MR, Sauerwein HP, Groener JE, Aerts JM, Ackermans MT, Glatz JF, Fliers E, Serlie MJ: **Gender-related differences in the metabolic response to fasting.** J Clin Endocrinol Metab 2007, **92**(9):3646-3652.

236. Ter Horst KW, Gilijamse PW, de Weijer BA, Kilicarslan M, Ackermans MT, Nederveen AJ, Nieuwdorp M, Romijn JA, Serlie MJ: **Sexual Dimorphism in Hepatic, Adipose Tissue, and Peripheral Tissue Insulin Sensitivity in Obese Humans.** Frontiers in endocrinology 2015, **6**:182.

237. Basu R, Dalla Man C, Campioni M, Basu A, Klee G, Toffolo G, Cobelli C, Rizza RA: **Effects of age and sex on postprandial glucose metabolism: differences in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction.** Diabetes 2006, **55**(7):2001-2014.

238. Varlamov O, Bethea CL, Roberts CT, Jr.: **Sex-specific differences in lipid and glucose metabolism.** Frontiers in endocrinology 2014, **5**:241.

239. Pietropaoli AP, Glance LG, Oakes D, Fisher SG: **Gender differences in mortality in patients with severe sepsis or septic shock.** Gender medicine 2010, **7**(5):422-437.

240. Valentin A, Jordan B, Lang T, Hiesmayr M, Metnitz PG: **Gender-related differences in intensive care: a multiple-center cohort study of therapeutic interventions and outcome in critically ill patients.** Crit Care Med 2003, **31**(7):1901-1907.

241. Eachempati SR, Hydo L, Barie PS: **Gender-based differences in outcome in patients with sepsis.** Arch Surg 1999, **134**(12):1342-1347.

242. Epstein SK, Vuong V: **Lack of influence of gender on outcomes of mechanically ventilated medical ICU patients.** Chest 1999, **116**(3):732-739.

243. Reinikainen M, Niskanen M, Uusaro A, Ruokonen E: **Impact of gender on treatment and outcome of ICU patients.** Acta Anaesthesiol Scand 2005, **49**(7):984-990.

244. Schroder J, Kahlke V, Staubach KH, Zabel P, Stuber F: **Gender differences in human sepsis.** Arch Surg 1998, **133**(11):1200-1205.

245. Dickson JL, LeCompte AJ, Floyd RP, Chase JG, Lynn A, Shaw GM: **Development and optimisation of stochastic targeted (STAR) glycaemic control for pre-term infants in neonatal intensive care.** Biomed Signal Process Control 2013, **8**.

246. Uyttendaele V, Dickson JL, Morton SE, Shaw GM, Desaive T, Chase JG: **Changes in Identified, Model-based Insulin Sensitivity can be used to Improve Risk and Variability Forecasting in Glycaemic Control.** IFAC-PapersOnline 2018, **51**(15):311-316.

247. Dickson J, Floyd RP, Le Compte A, Fisk L, Chase JG: **External validation and sub-cohort analysis of stochastic forecasting models in NICU cohorts.** Biomedical Signal Processing and Control 2013, **8**(4):409-419.

248. Uyttendaele V, Dickson J, Shaw G, Desaive T, Chase JG: **Improved 3D Stochastic Modelling of Insulin Sensitivity Variability for Improved Glycaemic Control.** IFAC-PapersOnLine 2018.

249. Uyttendaele V, Knopp JL, Shaw GM, Desaive T, Chase JG: **3D Stochastic Modelling of Insulin Sensitivity in STAR: Virtual trials analysis.** IFAC-PapersOnline 2018, **51**(27):128-133.

250. Sheather SJ: **Density Estimation.** In Statistical Science. Volume 19: Institute of Mathematical Statistics; 2004.[JSTOR (Series Editor)].

251. Scott DW: **Multivariate Density Estimation and Visualization.** In Handbook of Computational Statistics: Concepts and Methods. Edited by Gentle JE, Härdle WK, Mori Y. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012: 549-569.

252. James G, Witten D, Hastie T, Tibshirani R: **Resampling Methods.** In An Introduction to Statistical Learning. Volume 103: Springer, New York, NY; 2013: 175-201.[vol Springer Texts in Statistics].

253. Villet S, Chiolero RL, Bollmann MD, Revelly JP, Cayeux RNM, Delarue J, Berger MM: **Negative impact of hypocaloric feeding and energy balance on clinical outcome in ICU patients.** Clin Nutr 2005, **24**(4):502-509.

254. Heyland DK, Cahill N, Day AG: **Optimal amount of calories for critically ill patients: depends on how you slice the cake!** Crit Care Med 2011, **39**(12):2619-2626.

255. Rosa C, Donado JH, Restrepo AH, Quintero AM, Gonzalez LG, Saldarriaga NE: **Strict glycaemic control in patients hospitalised in a mixed medical and surgical intensive care unit: a randomised clinical trial.** Crit Care 2008, **12**.

256. Davidson S, Pretty C, Uyttendaele V, Knopp JL, Desaive T, Chase JG: **Multi-input stochastic prediction of insulin sensitivity for tight glycaemic control using insulin sensitivity and blood glucose data.** Comput Methods Programs Biomed 2019, **182**.

257. Davidson S, Uyttendaele V, Pretty C, Knopp JL, Desaive T, Chase JG: **Virtual patient trials of a multi-input stochastic model for tight glycaemic control using insulin sensitivity and blood glucose data.** Biomedical Signal Processing and Control 2020.

258. Dickson J, Chase JG: **Clinical Compliance in Personalised Model-based Medical Decision Support: Do computers and interfaces yield better compliance?** IFAC-PapersOnLine 2019, **51**(34):341-346.

259. Uyttendaele V, Knopp JL, Pirotte M, Morimont P, Lambermont B, Shaw G, Desaive T, Chase JG: **Preliminary results from the STAR-Liège clinical trial: Virtual trials, safety, performance, and compliance analysis.** IFAC-PapersOnLine 2018, **51**(27):355-360.

260. Uyttendaele V, Knopp JL, Pirotte M, Morimont P, Lambermont B, Shaw GM, Desaive T, Chase JG: **Translating A Risk-Based Glycaemic Control Framework for Critically Ill Patients: STAR-Liège.** IFAC-PapersOnline 2020:6-pages.

261. Finfer S, Chittock D, Li Y, Foster D, Dhingra V, Bellomo R, Cook D, Dodek P, Hebert P, Henderson W, et al: **Intensive versus conventional glucose control in critically ill patients with traumatic brain injury: long-term follow-up of a subgroup of patients from the NICE-SUGAR study.** Intensive Care Med 2015, **41**(6):1037-1047.

262. Gunst J, Berghe G: **Blood glucose control in the ICU: don't throw out the baby with the bathwater!** Intensive Care Med 2016, **42**.

263. Chase JG, Knopp JL: **Controlling Nutrition in the ICU and Other Heresies: An analysis, proof and review of the need to modulate nutrition for safe, effective glycemic control.** IFAC-PapersOnline 2020.

264. Penning S: **Tight Glycaemic Control : Model-based methods to answer critical questions about this controversial therapy.** PhD University of Liège; 2014.

265. Fléchet M: **Safety, Performance and Compliance: Clinical and in Silico Evaluation and Optimisation of Glycemic Control in Three Countries.** ME University of Liège, Biomedical Engineering; 2014.

266. Stewart KW, Thomas F, Pretty C, Chase JG, Shaw GM: **How should we interpret retrospective blood glucose measurements? Sampling and Interpolation.** IFAC-PapersOnline 2017, **50**(1):874-879.

267. Penning S, Lambermont B, Desaive T, Pretty C, Chase JG: **In silico assessment of a computarized model-based glycaemic control approach in a Belgian medical intensive care unit.** IFAC Proceedings Volume 2014, **47**(3):9913-9918.

268. Meijering S, Corstjens AM, Tulleken JE, Meertens JH, Zijlstra JG, Ligtenberg JJ: **Towards a feasible algorithm for tight glycaemic control in critically ill patients: a systematic review of the literature.** Crit Care 2006, **10**(1):R19.

269. Uyttendaele V, Knopp JL, Shaw GM, Desaive T, Chase JG: **Risk and Reward: Extending stochastic glycaemic control intervals to reduce workload.** Biomed Eng Online 2020.

270. Goldberg PA, Siegel MD, Russell RR, Sherwin RS, Halickman JI, Cooper DA, Dziura JD, Inzucchi SE: **Experience with the continuous glucose monitoring system in a medical intensive care unit.** Diabetes Technol Ther 2004, **6**(3):339-347.

271. Holzinger U, Warszawska J, Kitzberger R, Wewalka M, Miehsler W, Herkner H, Madl C: **Real-time continuous glucose monitoring in critically ill patients: a prospective randomized trial.** Diabetes Care 2010, **33**(3):467-472.

272. Preiser JC, Chase JG, Hovorka R, Joseph JI, Krinsley JS, De Block C, Desaive T, Foubert L, Kalfon P, Pielmeier U, et al: **Glucose Control in the ICU: A Continuing Story.** J Diabetes Sci Technol 2016, **10**(6):1372-1381.

273. Preiser JC, Lheureux O, Thooft A, Brimioulle S, Goldstein J, Vincent JL: **Near-Continuous Glucose Monitoring Makes Glycemic Control Safer in ICU Patients.** Crit Care Med 2018, **46**(8):1224-1229.

274. Cahill NE, Dhaliwal R, Day AG, Jiang X, Heyland DK: **Nutrition therapy in the critical care setting: what is "best achievable" practice? An international multicenter observational study.** Crit Care Med 2010, **38**(2):395-401.

275. Zhou T, Dickson JL, Shaw GM, Chase JG: **Continuous Glucose Monitoring Measures Can Be Used for Glycemic Control in the ICU: An In-Silico Study.** J Diabetes Sci Technol 2018, **12**(1):7-19.

276. Krinsley JS, Chase JG, Gunst J, Martensson J, Schultz MJ, Taccone FS, Wernerman J, Bohe J, De Block C, Desaive T, et al: **Continuous glucose monitoring in the ICU: clinical considerations and consensus.** Crit Care 2017, **21**(1):197.

277. Ormsbee J, Knopp JL, Chase JG: **Estimating (unidentifiable) enhanced EGP in glycaemic control modelling: Dancing with minions of the Dark Lord.** IFAC-PapersOnline 2020:6-pages.

278. Anane Y, Benyo B, Chase JG: **Clinical application scenarios to handle insulin resistance and high endogenous glucose production for intensive care patients.** IFAC-PapersOnline 2020:6-pages.

279. Casaer MP, Wilmer A, Hermans G, Wouters PJ, Mesotten D, Van den Berghe G: **Role of disease and macronutrient dose in the randomized controlled EPaNIC trial: a post hoc analysis.** Am J Respir Crit Care Med 2013, **187**(3):247-255.

280. Singer P, Anbar R, Cohen J, Shapiro H, Shalita-Chesner M, Lev S, Grozovski E, Theilla M, Frishman S, Madar Z: **The tight calorie control study (TICACOS): a prospective, randomized,**

controlled pilot study of nutritional support in critically ill patients. Intensive Care Med 2011, **37**(4):601-609.

281. Gunst J: **Recovery from critical illness-induced organ failure: the role of autophagy.** Crit Care 2017, **21**(1):209.

282. Gunst J, Van den Berghe G: **Intensive Care Nutrition and Post-Intensive Care Recovery.** Crit Care Clin 2018, **34**(4):573-583.

283. Chase JG, Shaw GM, Lin J, Doran CV, Hann C, Robertson MB, Browne PM, Lotz T, Wake GC, Broughton B: **Adaptive bolus-based targeted glucose regulation of hyperglycaemia in critical care.** Med Eng Phys 2005, **27**(1):1-11.

284. Wong XW, Chase JG, Shaw GM, Hann CE, Lotz T, Lin J, Singh-Levett I, Hollingsworth LJ, Wong OS, Andreassen S: **Model predictive glycaemic regulation in critical illness using insulin and nutrition input: a pilot study.** Med Eng Phys 2006, **28**(7):665-681.

285. Doran CV, Hudson NH, Moorhead KT, Chase JG, Shaw GM, Hann CE: **Derivative weighted active insulin control modelling and clinical trials for ICU patients.** Med Eng Phys 2004, **26**(10):855-866.

286. Lehmann ED, Deutsch T: **Insulin dosage adjustment in diabetes.** Journal of biomedical engineering 1992, **14**(3):243-249.

287. Hann CE, Chase JG, Ypma MF, Elfring J, Mohd Nor N, Lawrence P, Shaw GM: **The impact of parameter identification methods on drug therapy control in an intensive care unit.** Open Med Inform J 2008, **2**:92-104.