UNIVERSITY OF TWENTE.

zgt
Topzorg voor uw levenskwaliteit

Reality-centered AI for tighter in-hospital glucose control

A master's thesis for degree of Technical Medicine

by Robin Rijnhout

December 2, 2024

Enschede

Supervisors:

Prof. Dr. G. D. Laverman | Chairman/Clinical supervisor

Dr. Ir. B.J.F. van Beijnum | Technical supervisor

T. Urgert, MSc | Daily supervisor

Dr. M. Groenier | Process supervisor

Dr. K.D.R. Kappert | External member

Preface

Presented here is my master's thesis, on which I have spent nearly a year working. It's titled "Reality-centered AI for tighter in-hospital glucose control". I know AI is a real buzzword, and to be frank, that is also part of the reason why I used it so prominently in my title. But the part of the title that I would like the focus to be on is rather "Reality-centered". I am glad to use what I have learned in the past 6 years as a Technical Physician in training to bring technology, AI in this case, to the clinic and truly try to see what is needed to help persons with diabetes using technology. I hope this outcome does indeed have an impact or at least inspires others to have an impact in a similar way.

I would not have been able to do this if it were not for a lot of people around me. Firstly, I would like to express my gratitude to my graduation committee. Thank you, Goos, for the endless opportunities you have provided me in and around my graduation. With all your knowledge and enthusiasm, you are of huge value to me and a lot of other Technical Medicine students. Bert-Jan, for always keeping an eye out for me and the interesting discussions we have had over this year. You have challenged me multiple times to think critically about this research, which I might have found frustrating at times, but in the end, I am very thankful for it. Thomas, for being the bridge (a true Technical Physician) between the technology and the clinic. Also, for putting the whole process into perspective and, with that, comforting me in times when I was not so sure. Marleen, a well-meant thank you for all the conversations we have had about my professional development. I know it must not be easy for a teacher in professional behavior to spark interest in this topic in all these students, but with me, you have succeeded. Thanks to you, I have the toolkit to be the lifelong learner the program wants me to be. And a thank you to Kilian for reading this big chunk of a thesis (for which I apologize; as a lifelong learner, I acknowledge that concise writing will be my next goal) and bringing your perspective to the table.

A big thank you to my colleagues at the ZGT and my peer review groups at the University. All the fun, stress, breaks, and conversations we shared have been a huge part of this year for me. A special thanks to Stennie, Sacha, Eclaire, Chiara, and Anouk, who shared their wisdom but also fun and enthusiasm with me. To all my friends, who are too much to name, who have supported me throughout this year and brought the necessary distraction.

To my dear girlfriend Femke, who has been there and unconditionally supported me at every moment of this thesis and more. For always reminding me how proud you are of me and how proud I should be of myself. And at last, to my parents and my sister. I cannot express enough gratitude for everything you have done for me and all the opportunities you gave me in my entire life, which ultimately led to this milestone. Thank you all.

Table of contents

Preface **1**

List of abbreviations **4**

1. Report introduction **5**

2. Chapter 1: Patient-specific prediction models to improve inpatient glucose regulation **7**

2.1 Abstract *7*

2.2 Introduction *8*

2.3 Methods *12*

2.4 Results *21*

2.5 Discussion *35*

2.6 Conclusion *41*

3. Chapter 2: Challenges for doctors in glucose regulation in hospitalized patients: a qualitative insight from medical doctors working in the department of internal medicine **42**

3.1 Abstract *42*

3.2 Introduction *42*

3.3 Methods *43*

3.4 Results *44*

3.5 Discussion *49*

3.6 Conclusion *51*

4. Chapter 3: The patient's story: two patient personas **52**

4.1 Abstract *52*

4.2 Introduction *52*

4.3 Methods *53*

4.4 Results *54*

4.5 Discussion *57*

4.6 Conclusion *58*

5. Final report conclusions & a proposal for a simple decision-support system **59**

5.1 Conclusions *59*

5.2 Proposal for a simple decision-support system *60*

Bibliography **63**

Appendix: **67**

Literature ARIMA(X) modelling for BG prediction *67*

Interview questions *70*
Introduction: 70
Questions: 70

List of abbreviations

1. BG = Blood glucose
2. LOS = length of stay
3. CGM = continuous glucose monitoring
4. CSII = continuous subcutaneous insulin infusion
5. AID = automated insulin delivery system
6. T1D = type 1 diabetes
7. T2D = type 2 diabetes
8. ARIMA(X) = autoregressive integrated moving average (with exogenous variables)
9. PH = prediction horizon
10. BIC = Bayesian information criterion
11. AIC = Aikake information criterion
12. RMSE = root mean square error
13. MAE = mean absolute error
14. MAPE = mean absolute percentage error
15. MARD = mean absolute relative distance
16. CEGA = Clarke's error grid analysis
17. ADF = augmented Dickey-Fuller test
18. KPSS = Kwiatkowski-Phillips-Schmidt-Shin
19. ANOVA = Analysis of variance
20. MPC = model predictive control
21. EHR = Electronic Health Record
22. TDD = total daily dose

1. Report introduction

Persons with diabetes have a higher chance of hospitalization than persons without diabetes. Approximately 25% to 40% of all hospitalized patients suffer from pre-existing diabetes (Pérez et al., 2020). In the ZGT, in 2023, persons with diabetes accounted for approximately 10% of all admissions. As diabetes rates continue to rise globally, the total number of hospital admissions will likewise increase (Standl et al., 2019). Glucose regulation in the hospital setting is more challenging than in a domestic environment. The context significantly differs from that of a domestic environment. New factors introduced that affect glucose regulation include the level of illness, different nutritional patterns and modalities, induction of medication, and reduced monitoring frequency of blood glucose (BG) values. As a result of these factors introduced in a hospital setting, physicians must take over BG regulation treatment in some cases (Moghissi et al., 2009; Pérez et al., 2020).

BG regulation treatment efforts pose a high risk of hypoglycemia, an independent risk factor for mortality. To prevent hypoglycemia, conservative glucose management is often triggered by physicians. This results in more frequently tolerated hyperglycemia among hospitalized patients (Moghissi et al., 2009a; Pérez et al., 2020a). Observational evidence links hyperglycemia in hospitalized patients to increased mortality, longer length of stay (LOS), and slower wound healing (Pérez et al., 2020a). Improved glucose control is expected to decrease these negative effects. Moreover, the likelihood of systemic infection diminishes (Moghissi et al., 2009a).

Current hospital workflows regarding BG regulation have limitations. Workflow plays an important role in managing the risk of hypo- and hyperglycemia. Failure of a physician to adjust glycemic therapy according to blood glucose patterns can increase the risk of experiencing both hypo- and hyperglycemia(Moghissi et al., 2009b). This can, in part, be due to a lack of BG values documented in the electronic health record (EHR) (Rousseau et al., 2014). Insulin and nutrition are often hard to find or undocumented in the EHR. As diabetes is frequently not the primary reason for admission, there can be poor communication in times of transfer. The communication, availability of data, and adjusting medication according to that data are requirements to facilitate a workflow that ensures tight BG regulation (Moghissi et al., 2009b; Rousseau et al., 2014)

It becomes clear that there is a need for tight glucose regulation for inpatients while reducing the risk of hypoglycemia. The use of technologies like continuous glucose monitoring (CGM), continuous subcutaneous insulin infusion (CSII), and/or a combination of CGM and CSII: automated insulin delivery (AID) can lead to a decrease in episodes of hyper-and hypoglycemia outside of the hospital environment (Umpierrez & Klonoff, 2018). This effect arises from higher BG measurement frequencies, enabling more effective insulin treatment via CSII, or AID. Nevertheless, glycemic control with CGM does not show significant improvement when compared to point-of-care

measurements in hospitalized patients (Spanakis et al., 2022). This indicates that applying only CGM in a hospital environment is not enough to increase insight into a person with diabetes's data.

Regular insights into BG data are essential for the effective detection of hyper- or hypoglycemic events with CGM. For these insights, the nurses can make rounds to gather CGM data from the admitted patients, but this is labor-intensive in an already tight schedule for nurses. A study has been performed to transfer CGM data automatically to a central nursing station to prevent hypoglycemia (Spanakis et al., 2018). The study reported successful prevention of clinically relevant and severe hypoglycemia in all five participants. The telemetry system also has the potential to treat hyperglycemia. However, Spinakis et al. only proposed a simplified hypoglycemia prevention protocol to combine with their telemetry system. They propose that to prevent hyperglycemia, nursing staff must enhance their understanding of CGM data interpretation. To prevent hypo- or hyperglycemia, the success of this system largely depends on the correct interpretation of data by nursing staff.

Additional challenges arise with the implementation of CGM and AID in the hospital. The length of an admission period and the changing routine make it hard for AID algorithms to regulate glucose. In another scenario, due to illness and differentiating levels of consciousness, patients may be deemed incompetent to manage their glucose regulation with CGM and AID, so AID is stopped. AID may also be stopped upon admission due to diabetic ketoacidosis, before and during extended surgical procedures, or if CGM and AID supplies are unavailable in the hospital. Should CGM and AID remain in use, physicians and nurses must have the essential skills to manage these technologies. However, this expertise is frequently lacking (Umpierrez & Klonoff, 2018).

Frequent monitoring, insight into essential glucose and patient data, and a sufficient workflow are needed to satisfy the need for goog glycemic control for inpatients while reducing the risk of hypoglycemia. To combine these features, a decision-support system is designed. This report focuses on insight into problems with glucose regulation in the hospital and the development of a decision-support system tailored towards in-hospital care. The overall objective of this report is:

To provide a comprehensive analysis of the challenges related to glucose regulation in hospitalized patients and to develop a decision-support system to improve clinical management and optimize patient outcomes.

The report is divided into three chapters to achieve this goal. The main chapter is on the development of models for BG prediction inside of a decision-support system for in-hospital glucose regulation. Two smaller chapters introduce extra features for the system and give more context to how this system should be implemented in hospital care. At the end of the report, a general conclusion is drawn.

2. Chapter 1: Patient-specific prediction models to improve inpatient glucose regulation

2.1 Abstract

Introduction: BG prediction aids in decision-support for BG regulation. This chapter explores patient-specific autoregressive integrated moving average models with exogenous variables (ARIMA(X)). Literature reveals that ARIMA(X) models have not been evaluated in a hospital context. Besides this, there is limited attention to the clinical relevance regarding safety and usability in literature. This study assesses patient-specific ARIMA(X) models in an adaptive identification framework, utilizing available data and considering diabetes's time-variant dynamics.

Methods: A retrospective cohort study was performed. Quantitative, observational, and secondary CGM, insulin, and carbohydrate data were used from the DIABASE database. Data was extracted from hospitalized patients, including both data from during and right before hospitalization. CGM with >20% missing data was excluded. To impute the left over CGM with missing data, seasonal decomposition and interpolation were performed. Model identification was conducted to optimize ARIMA(X) parameters. Predictions were made for PH = {30 min, 60 min, 120 min}. An adaptive identification algorithm was applied for continuously changing model orders. Testing was performed using time series cross-validation. Performance metrics were chosen based on clinical relevance: MAE, RMSE, MAPE, CEGA, amount of training data needed for (accurate) prediction, and computational time.

Results: This study contained a group of 15 patients (mean age 56.33 years, n = 10 female), with some patients having multiple admissions. The MAE was, respectively, 0.65 (±0.21), 1.17 (±0.23), and 1.88 (±0.24) mmol/l for PH = 30 min, 60 min, and 120 min. The RMSE was, respectively, 0.87 (±0.27), 1.52 (±0.27), and 2.41 (±0.25) mmol/l for PH = 30 min, 60 min, and 120 min. The models trained and tested with intra-extended data reported totals of 100%, 99.87%, and 99.42% in the A+B zones of the CEG for PH = 30 min, 60 min, and 120 min. No dataset satisfies the set threshold for all PH when compared to the MAPE threshold of 10%. For PH = 30 min., the models trained and tested on intra-BG, pre-BG, cross-BG, and intra-extended data complied with the threshold. Predictions were computed quickly and accurately with little training data available.

Discussion: Cross-BG data showed the most promising predictions that satisfied the different criteria and benchmarks of clinical relevance across various PHs. The satisfaction of the criteria and benchmarks for clinical relevance depended heavily on the different PHs. The adaptive identification algorithm could be studied more elaborately on purposefully collected and simulated patient data. The same study could

also analyze the models' sensitivity to insulin and carbohydrate inputs. For this study to have direct implications, a study could be performed on the use of the model as a decision-support system either only as a 30-minute-ahead alarm or in combination with the bolus algorithm of Pérez et al. under strict expert supervision.

2.2 Introduction

Decision-support systems for glucose regulation predict glucose to estimate the amount of insulin, either bolus, basal, or both, or the amount of carbohydrates needed to keep the BG values of a patient within the desired range. The predictions can help induce preventive actions when predicting hyper- or hypoglycemic events (Prendin et al., 2021). BG prediction models can either be physiological prediction models, data-driven models, or a hybrid version of those options (Oviedo et al., 2017). This chapter focuses on a data-driven approach, namely, autoregressive integrated moving average models with exogenous variables (ARIMA(X)).

Research indicates that various approaches have been implemented in ARIMA(X) models and similar frameworks over time. These models have been utilized with datasets from real-life settings. The findings from these diverse studies are compiled in Table 1, these results will be used as comparison later in this report. Comparing and choosing the best-performing models is challenging due to the distinct data sets utilized in these studies. To determine the most suitable model for this study, a thorough examination of methodologies presented in the literature is conducted. A comprehensive outline of the reviewed studies, including their rationales, methods, and results, can be found in the Appendix of this report. Eventually, for this study's context, patient-specific ARIMA(X) models within an adaptive identification framework were chosen, leveraging all pertinent data while considering the time-variant dynamics of diabetes. This approach parallels the techniques employed by Yang et al. (2019). It is anticipated that the adaptability of this method makes the models appropriate for an exploration of what works well in the novel context this study introduces.

Table 1 Summary of results from literature for context and for comparison with this study's method. The lowest RMSE for relevant PH is in bold. The highest Clarke error grid analysis (CEGA) is not shown in bold, as only one study contained this information.

Study	**Model**	**PH (minutes)**	**RMSE (mmol/l)**	**MAE (mmol/l)**	**MAPE(%)**
Phadke et al. 2020	*ARIMA(2,1,0)*	*Libre Pro Dataset*			
		15	0.39	0.28	3.98
		30	0.88	0.64	9.37
		45	1.29	0.97	13.87
		Ohio T1DM Dataset			
		15	0.73	0.44	8.21
		30	1.29	0.84	11.47

		45	1.79	1.21	16.62
Predin et al. 2021		Individual order			
	AR	30	1.26 [1.04-1.64]	-	-
	ARMA		1.25 [1.12-1.69]	-	-
	ARIMA		1.23 [1.1-1.60]	-	-
		Individual order 30-minute specific			
	AR	30	1.27 [1.1-1.6]	-	-
	ARMA		1.27 [1.14-1.66]	-	-
	ARIMA		1.24 [1.11-1.63]	-	-
		Individual order Day and Night			
	AR	30	1.35 [1.15-1.68]	-	-
	ARMA		1.35 [1.18-1.68]	-	-
	ARIMA		1.28 [1.14-1.65]	-	-
		Regularized			
	AR	30	1.29 [1.1-1.72]	-	-
		RLS			
	AR	30	1.52 [1.37-1.88]	-	-
Prendin et al. 2022		Ohio T1DM Dataset			
	ARIMA	30	1.09 [1.02-1.14]	-	-
		45	1.5 [1.33-1.59]	-	-
		60	1.87 [1.66-1.95]	-	-
		75	2.17 [1.8-2.31]	-	-
	ARIMAX	30	1.04 [0.96-1.11]	-	-
		45	1.47 [1.28-1.50]	-	-
		60	1.71 [1.63-1.77]	-	-
		75	1.92 [1.74-2.17]	-	-
		CTR3 Dataset			

	ARIMA	30	1.17 [1.11-1.38]	-	-
		45	1.63 [1.52-1.85]	-	-
		60	1.97 [1.92-2.25]	-	-
		75	2.45 [2.19-2.55]	-	-
	ARIMAX	30	1.16 [0.99-1.3]	-	-
		45	1.56 [1.35-1.81]	-	-
		60	1.87 [1.59-2.25]	-	-
		75	2.22 [1.74-2.41]	-	-
Mohebbi et al. 2020	ARIMA (With 7 days of trainingdata)	15	0.68 [±0.05]	0.46 [±0.03]	-
		30	1.22 [±0.08]	0.86 [±0.06]	-
		45	1.70 [±0.12]	1.23 [±0.08]	-
		60	2.14 [±0.15]	1.57 [±0.11]	-
		90	2.83 [±0.21]	2.12 [±0.15]	-
Sawaryn 2020	AR (optimal training set sizes, mean 9.6 days ±3.01, and parameters, mean AR: 12 ± 2 per patient)	15	0.33 [±0.08]	-	-
		30	0.8 [±0.15]	-	-
		60	1.44 [±0.2]	-	-
		120	1.95 [±-.21]	-	-
		180	2.16 [±0.2]	-	-
	ARMAX (optimal training set sizes, mean 10.4 days ± 1.62, and parameters, mean AR & X: 7.6 ± 0.8, MA: 8.2	15	0.32 [±0.08]	-	-
		30	**0.76 [±0.16}**	-	-
		60	**1.30 [±0.21]**	-	-
		120	**1.64 [±0.24]**	-	-
		180	1.74 [±0.25]	-	-

	±0.4, Nk: 1 per patient)					

Study	**Mean values (%) in CEGA zones at PH = 120 min.**					
Sawaryn 2020	Models	A	B	C	D	E
	AR	67.9 [±6.81]	31.0 [±7.22]	0.07 [±0.13]	1.04 [±0.66]	0.0 [±0.0]
	ARMAX	74.6 [±7.58]	24.8 [±7.56]	0.03 [±0.07]	1.54 [±0.54]	0.0 [±0.0]

Although based on a limited literature review, this analysis revealed several gaps in relation to the objectives of this study. The literature shows the potential of ARIMA(X) or models alike to predict BG values in persons with diabetes. However, it lacks evidence that this is also possible in a hospital scenario. Besides this, there is limited attention to the clinical relevance of model performance in terms of safety and usability. Clinically relevant prediction is defined as accurate, safe, and usable in a hospital context. Safety and usability in a certain context are lacking in most literature. If a prediction satisfies the requirements for those three categories, it can lead to tighter glucose regulation while minimizing the risk of hypoglycemia.

To fill this gap, this research has the following objective:

This study aims to assess a clinically relevant BG prediction by developing different patient-specific ARIMA(X) models in an adaptive identification framework for persons with diabetes in non-ICU hospital wards.

To obtain the research objective, a set of sub-questions needs to be answered:

1. Which input signals result in the most clinically relevant BG prediction model for prediction horizon of 30 min, 60 min, and 120 min?
2. What amount of training data results in the most clinically relevant BG prediction model for prediction horizon of 30 min, 60 min, and 120 min?
3. What is the added value of a patient-specific adaptive identification time series modeling approach for BG prediction?

To answer these questions, this study evaluates different patient-specific ARIMA(X) models in an adaptive identification framework.

The existing literature in different contexts functions as a starting point for this research. Based on this literature, it is expected that in that clinically relevant patient-specific models can be developed to make predictions for hospitalized patients at the prediction horizons (PH) of 30 min, 60 min, and 120 min. Sawaryn et al. showed that predictions with AR and ARMAX models can be performed safely and accurately in persons with type

2 diabetes (T2D) for PH = 30 min, 60 min, and 120 min, which means that predictions almost exclusively fell in zones A & B of the Clarke error grid. Besides this, unlike non-linear alternatives, Sawaryn et al. and Mohebbi et al. showed that autoregressive models need little training data to make predictions, which shows the usability of the models in various contexts (Mohebbi et al., 2020; Sawaryn, 2020).

This report summarizes the process of obtaining the research objective and questions. The method section elaborates on data pre-processing, model structure and development, and assessment of clinical relevance, after which the results and discussion are presented for this chapter.

2.3 Methods

To reach the objective of developing and assessing a clinically relevant BG prediction model for people with diabetes in non-ICU hospital wards, a retrospective cohort study was performed. Three main phases were distinguished in the method. First, data was selected and pre-processed. Second models were developed and tested. Lastly, the results from the model testing were analyzed. These phases are discussed in more detail in this section. Figure 1 is a summary of the complete method section.

Quantitative, observational, and secondary data were used from the DIABASE database (ZGT, Almelo). This database contained CGM, AID, and other relevant medical data from persons with type 1 diabetes (T1D) and T2D with CGM and/or AID. Data was extracted from hospitalized patients, including both data during and right before hospitalization. A set of inclusion and exclusion criteria was set to obtain effective data.

2.3.1 Inclusion criteria

- The patient had given informed consent for inclusion into the DIABASE database.
- The patient was ≥ 18 years old.
- The patient had T1D or T2D.
- CGM data was consecutively available for ≥ 24 hours during hospitalization

2.3.2 Exclusion criteria

- The patient had retrieved their informed consent for the DIABASE study during this research.
- The patient was hospitalized in an ICU-ward.

The number of patients with sufficient and suitable data based on the research goal and in- and exclusion criteria were selected at the beginning of data collection. The available CGM, AID, carbohydrate, time date, patient characteristics, and lab data were extracted from the database. Most models from the literature that predict BG showed good results using either consecutive BG data, insulin doses, and carbohydrate intake or a combination of those variables (Tsichlaki et al., 2022).

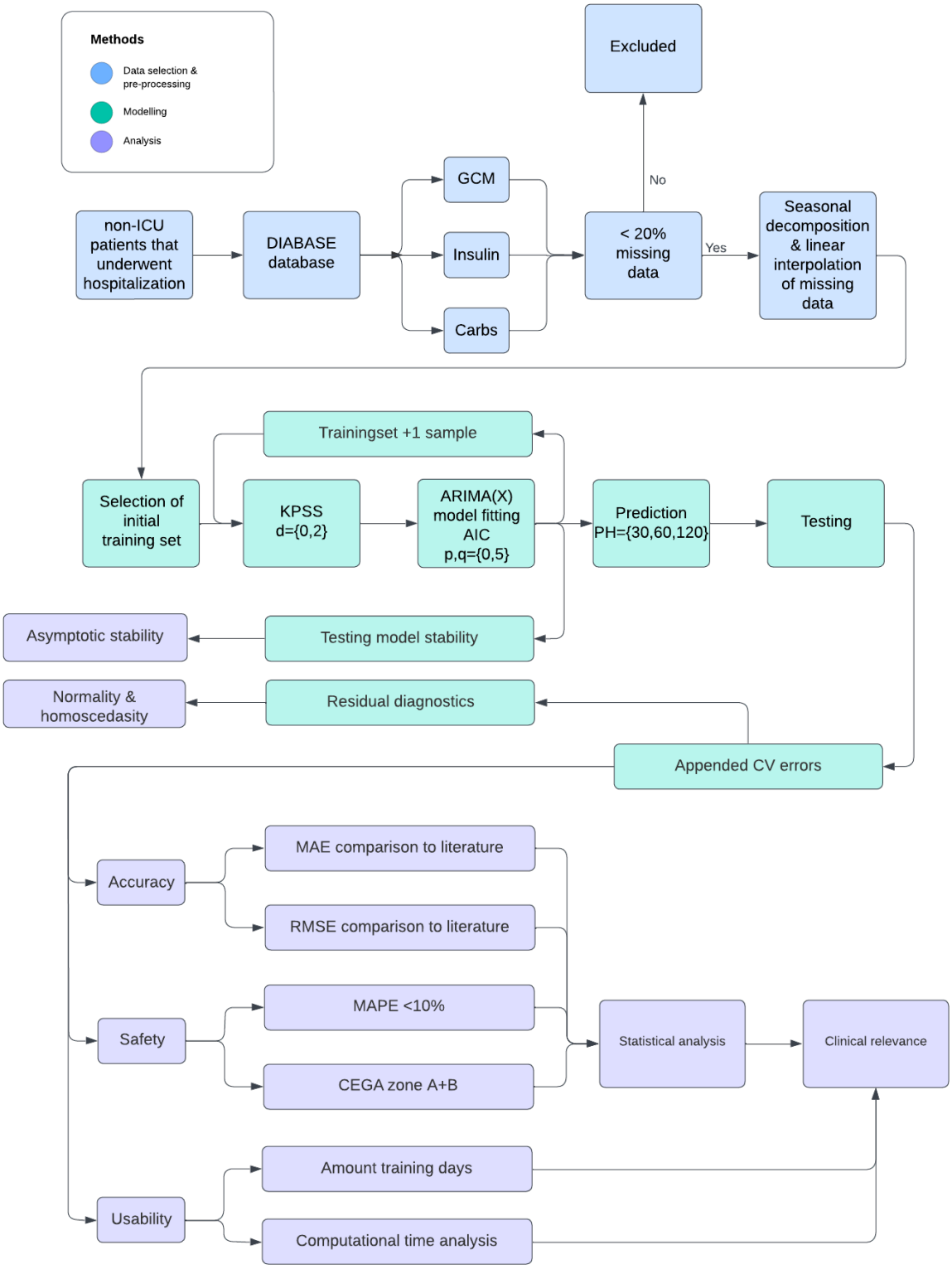


Figure 1 Flowchart summary of the method section of this study

2.3.3 Programming

The necessary programming for this study was performed with R 4.2 (Rstudio team, Boston, MA).

2.3.4 Input signal selection

Input signals for the models that were selected directly from the DIABASE dataset were CGM values, the amount of basal and bolus insulin, the amount of carbohydrate intake with their corresponding time data information.

CGM data was collected by different subcutaneous sensors, namely the Freestyle Libre Link and Medtronic sensor combined with an insulin pump. The Freestyle Libre Link collected a BG sample every 15 minutes, while the Medtronic sensor collected a BG sample every 5 minutes. To determine whether analysis of the BG prediction models should be performed separate for the different CGM sensors, a separate analysis was performed on the accuracy of the models using different sensors as input. The mean of the accuracy measures for all PH, which will be introduced later in this section, was compared between patients with a Freestyle Libre Link and Medtronic sensor. An independent t-test was performed on the mean absolute errors (MAE) of the two sensor groups. If $p < 0.05$, the group means were significantly different. If the different sensors yielded significantly different results, both sensor groups were split in the rest of the phases of development and analysis.

Insulin & carbohydrate data was retrieved from the Medtronic insulin pump. Insulin data contained basal and bolus inputs from the patient or from closed-loop pump settings. Carbohydrate data was dependent on patient input in the insulin pump. To answer sub question one, various input signal combinations were tested within the ARIMA(X) models to determine the most effective training method.

2.3.5 Missing data

Machine Learning studies lack a clear consensus on the threshold for including samples with missing data (Dong & Peng, 2013; Oluwaseye Joel et al., 2022). Consensus is lacking, for it is not only the amount of missing data but also the randomness of the missing data being important. In CGM, a study by Smith et al. concluded that large blocks of CGM were the most common missing data pattern. They argue for a cut-off of 30% missing data per day, as this results in a representative time in range (TIR) (Smith et al., 2023). In this study, a cut-off of 20% per day of recorded data was selected as it was observed during the setup of the study to yield good results in imputing the CGM.

Imputation of the missing data was performed by first seasonally decomposing the CGM data and applying imputation based on interpolation after. Out of linear-, mean- and moving-average imputation, this imputation method visually resembled the original signal the best.

2.3.6 Training and testing of the models

As a method for training and testing the models, time series cross-validation (CV) was applied. The model was trained and tested for PH = {30 min, 60 min, 120 min} for each step in the data samples. An example of how this works can be seen in Figure 2. The blue dots visualise the training data and the red dots the testing samples for PH = {30 min, 60 min, 120 min}. The size of the training set varied from 30 minutes to 17 days of training data, depending on available data from individual patients.

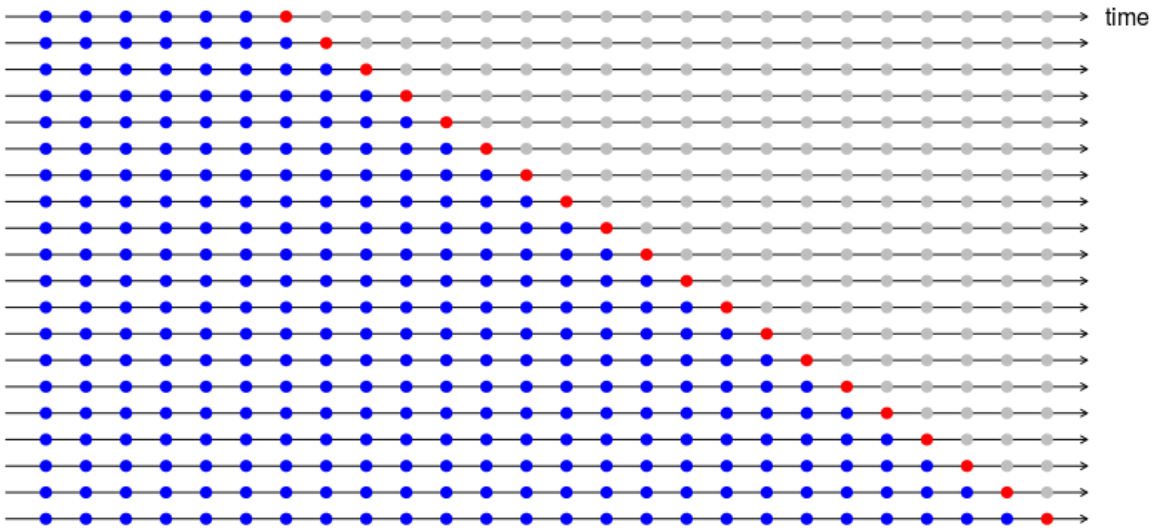


Figure 2 A visual representation of time series cross-validation where the blue dots resemble the training data and the red dot the test data. Source: https://robjhyndman.com/hyndsight/tscv/

2.3.7 ARIMA(X) model

An ARIMA(X) model was used to model the BG time series. An ARIMA(X) model consists of four parts, the AR, I, MA, and X part. A BG time series in a stationary form as input, which means a constant mean and variation over time, yields the best results when using an ARIMA(X) model. So, the first part of an ARIMAX model is the I term, which stands for integration, which helps to ensure this stationary form if a BG time series is not stationary on itself. To make the BG time series stationary, differencing steps can be taken. A differencing step replaces a BG value by the difference between each value and the previous value. Parameter d in an ARIMA(X) order representation shows the number of differencing steps needed to obtain a stationary BG time series.

Secondly, there is the AR part that only uses previous BG values to make a prediction. The AR part is defined by:

$$y'_t = c + \sum_{i}^{p} \phi_i y_{t-i} + \varepsilon_t \quad (1)$$

Where y'_t denotes the d times differenced BG time series at time t, y_{t-i} is a BG value at time $t - i$, ϕ_i is a parameter that describes how a previous BG value depends on the

prediction, ε_t is a white noise term which represents prediction error, and c is a constant. Depending on the amount of previous BG values that aid a BG prediction, p is a parameter that denotes the amount of previous BG terms incorporated in the AR part of the model.

Thirdly, the model includes an MA component that extends the model to incorporate previous prediction errors. Formula 1 extended with the MA part is defined by:

$$y'_t = c + \sum_{i}^{p} \phi_i y_{t-i} + \sum_{j}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (2)$$

where ε_{t-j} is a prediction error at time $t-j$, θ_j is a parameter that describes how a previous prediction error depends on the prediction. The number of previous prediction errors that aid a BG prediction is denoted by parameter q. Equation 2 represents an ARIMA model. The number of previous BG or prediction error values and the number differencing steps are often denoted as orders for an ARIMA model as (p, d, q).

Lastly, exogenous variables can be added, creating an ARIMAX model. Exogenous variables are variables that have a relation to the BG values, thus aiding prediction. Adding the X part to Formula 2, the ARIMAX model is defined by:

$$y'_t = c + \sum_{i}^{p} \phi_i y_{t-i} + \sum_{j}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t + \sum_{l}^{K} \beta_l X_{l,t} \quad (3)$$

where $X_{l,t}$ is the external input at time t, and $\beta_l$is a parameter that relates the exogenous variable to the predicted BG value. The number of previous values from the exogenous variables that are included is denoted by K.

An adaptive identification algorithms for (p,d,q) orders was employed for the ARIMA(X) model to account for time-varying data (Yang, 2019). First, an initial set of training data was selected. From this selected set the differencing step was determined for the CGM data by using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test as a test for stationarity. Yang et al. use the augmented Dickey-Fuller (ADF) test. However, the KPSS test has a null hypothesis of stationarity. This means the test selected differences if there was enough evidence to invalidate the stationary assumption(Kwiatkowski et al., 1992). ADF would select at least one difference unless the null hypothesis could be invalidated. If the p-value of the KPSS test was less than 0.05 for the original CGM series, parameter d is changed to 1. The process is repeated until stationarity for d = {0,…,2}. The best model fit on training data across various model orders is evaluated using the Aikake information criterion (AIC). The AIC balanced out the goodness of fit among the range of set parameters and incorporated a factor to avoid overfitting by also favoring a model with the smallest number of parameters. The model with the lowest

value for AIC is finally selected as the prediction model. A model is fitted to the data for orders p = {0,...,5} and q = {0,...,5}. These values were chosen as the maximum value in the analyzed literature for p was 30, and for q, 24. When tested among p = {0,..., 30} and q = {0,...,24}, for this dataset, orders did not exceed p = {0,...,5} and q = {0,...,5}, so the final order ranges were chosen as previously described. Coefficients for every set of orders were estimated using the conditional sum of squares of residuals to find initial values and to refine these initial values; maximum likelihood was applied. After the process of fitting the ARIMA model, the exogenous variables were fitted. This model fitting process is done with the *auto.arima()* function in R.

To evaluate the effectiveness of the adaptive algorithm on the dataset used in this study, an analysis of the changes in order over time was performed. The average changes in order over time for all participants were calculated. In addition, the share of the most occurring order was calculated.

Following the model fitting process, a prediction is made for PH = {30 min, 60 min, 120 min.} using the forecast() function in R. After this prediction, the training data was updated with the first sample of the last testing set, and a new model was fitted until the last data sample.

2.3.7 Input signal naming

To describe the different input signals for different models used in this study in an intuitive way that improves readability, a naming convention was developed. As introduced in the input signal section, CGM, insulin, and carbohydrate data from the DIABASE dataset were used for training and testing in different combinations. The models were trained and tested on data during, before, and combined between before and during admission. Prefixes and suffixes were developed based on the input signal and the time period of the data.

The prefix indicated the time period of data used for training and testing. The prefixes are reported in Table 2.

Table 2 Prefixes to indicate the time period from which the input signals used to make certain models stem

Prefix	Time period
Intra	Trained and tested with input signals from during the admission period
Pre	Trained and tested with input signals from before the admission period
Cross	Trained with input signals from before and during admission and tested on only on data during admission

The suffix indicated the set of input signals used to train the models. The suffixes are reported in Table 3.

Table 3 Suffixes to indicate the set of input signals used to make certain models

Suffix	Input signals used
BG	Only BG used as input signal into the ARIMA models
Extended	BG, insulin, and carbohydrates used as input signals into the ARIMAX models

An example of a combination of input signals from a certain time to show how the naming convention works is Cross-Extended, which indicates a model being trained on BG, insulin, and carbohydrate data from before admission and further trained and tested with BG, insulin, and carbohydrate data during admission.

2.3.8 Prediction horizon

As mentioned, the PHs are 30 min, 60 min, and 120 min for this study. Literature showed that treatment starting 20-30 minutes before hypo or hyperglycemia, is effective in resolving these episodes. It was shown that a PH = 30 min. is a good trade-off between limiting the error of the prediction outcome and effective prediction(Prendin et al., 2021). A larger prediction horizon leads to a deterioration in the accuracy of the prediction (Oviedo et al., 2017). However, PH = 60 min. and 120 min. were tested because of their possible impact on a decision-support system. In model predictive control for example, the BG PH should have the same length as the active time of a possible treatment option, to adjust the treatment accordingly (Crecil Dias et al., 2020). Fast-acting insulin has an action time of approximately 120 minutes (Summary of Product Characteristics: NovoRapid, n.d.). Which is why 120 min. was chosen as the largest prediction horizon.

2.3.9 Residuals check

The residuals of the model were checked on model assumptions. The residuals were checked on normal distribution and homoscedasticity. Normality was checked with a histogram and Shapiro-Wilk test; residuals were approximately normally distributed if $p>0.05$. Homoscedasticity was checked by plotting the residuals against the fitted values and doing a Ljung-box test; residuals were not significantly autocorrelated if $p < 0.05$. All diagnostics were performed for the residuals of different PH.

2.3.10 Stability analysis

All models were checked for asymptotic stability of the AR part. A stable ARIMA(X) model had all its poles and zeros inside of the unit circle. A short analysis was

performed on the different ranges of poles and zeros that resulted from the adaptive identification algorithm.

2.3.11 Performance metrics

Model performance was evaluated on clinical relevance in an inpatient context. That meant the model was evaluated based on accuracy, safety, and usability.

Model performance was compared to existing literature using the MAE and RMSE to assess accuracy.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y(t) - \hat{y}(t|t-PH)| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \hat{y}(t|t-PH)\right)^2} \quad (5)$$

Where PH is the prediction horizon, N is the length of a patient data portion of the test set, $y(t)$ is the current BG value and $\hat{y}(t|t-PH))$ is its PH-step-ahead prediction. While MAE represents a general picture of the accuracy of a model's predictions, RMSE penalizes large errors.

To assess the MAE and RMSE a literature benchmark was made. This benchmark consisted of the lowest and highest MAE and RMSE values from the studies reported in Table 1. Each prediction horizon considered in this study had its own benchmark, as can be seen in Table 4. An MAE & RMSE value was higher, within, or lower compared to the benchmark. In the final assessment, different contexts from the literature were considered.

Table 4 Literature benchmarks to make a comparison between the accuracy of this study's models and those from previous research.

Literature benchmark per metric	**PH = 30 min.**	**PH = 60 min.**	**PH = 120 min.**
MAE	0.64 – 0.86 mmol/l	1.57 mmol/l	-
RMSE	0.76 - 1.52 mmol/l	1.30 – 2.14 mmol/l	1.64 – 1.95 mmol/l

Safety was analyzed in two ways. Firstly, it was analyzed using the mean absolute percentage error (MAPE).

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y(t) - \hat{y}(t|t-PH)}{y(t)}\right| * 100 \quad (6)$$

This study considers a MAPE of 10% or lower as a safety criterion. This metric stems from CGM-sensor safety assessment[1].

To further test safety, Clarke error grid analysis (CEGA) was used as a second analysis. CEGA was originally developed to describe the accuracy of self-monitoring BG systems considering clinical relevance (Clarke et al., n.d.) The CEG consists of five regions, each having its own meaning:

- Region A: Predictions within 20% of reference
- Region B: Will not lead to inappropriate treatment
- Region C: Leading to unnecessary treatment
- Region D: Possibly dangerous failure to predict hypoglycemia or hyperglycemia
- Region E: Confuse treatment of hypoglycemia for hyperglycemia and the other way around

For a model to comply with the CEGA safety criterion, all predictions must fall into regions A & B of the CEG. Both the MAPE and CEGA safety criteria were assessed per PH.

For usability, the amount of input data needed for an accurate prediction, the minimal amount of data needed for prediction, and the computational time for a model to train and forecast were used as assessment. In an inpatient context, if patients only receive CGM during admission, it is important for the model to perform well with little data available. To assess this, the average RMSE across all PH of the model was analyzed for 1 day to 6 days of training data. An amount of training data was sufficient and thus useful if the RMSE at least fell within the average of the literature benchmark for RMSE across all PHs. Only RMSE was compared to the benchmark, as not enough MAE values from the literature were available to create a meaningful average. The averaged RMSE benchmark for this comparison was between 1.23 – 1.87 mmol/l. In addition, computational time analysis was performed. Computational time was defined as the time it took to pre-processing data, train a model, and predict BG values. Computational

[1] In research on CGM, mean absolute relative difference (MARD) is used to see whether the CGM sensor is feasible for insulin treatment decisions.

$$MARD = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{CGM_i - Comparison_i}{Comparison_i}\right| * 100 \quad (7)$$

A MARD of 10% is said to be the minimal accuracy a sensor should have. Further accuracy improvements do not contribute to better glycemic outcomes (Kovatchev et al., 2015). MAPE has the same mathematical description as MARD, the only difference being that a reference value is tested against a predicted value instead of sensor values. In the context of this study, the model had the purpose of leading to feasible insulin treatment decisions. Therefore, a MAPE of 10% or lower is considered as a safety criterion in this study.

time was tracked for different amounts of training data, from a single day to seven days. For this analysis, the model was trained on cross-extended data. These models contained the most variables and data amounts and are thus assumed to be the most complex and require the most computational time. Computational time was assessed on expert experience.

2.3.12 Statistical analysis

To give an answer to the first sub question, statistical analysis was performed to see whether the combination of input signals had a significant effect on the accuracy and safety metrics for all the PHs. First, a linear mixed effects model was made for every separate metric and PH. The fixed effect in the models was the input signal combination, and the individual patient admission was included as a random effect to take into account the variation between admission periods and overlapping admission periods in different input signal combination groups. ANOVA was performed to test the variance attributable to the fixed effect, input combination. If the input signal combination had a significant effect on the metric, post hoc tests were performed to see which pairwise differences were significant. Significance was determined at $p < 0.05$.

2.4 Results

2.4.1 Patient & admission characteristics

This study contained a group of 15 patients (mean age 56.33 years, n=10 female). Characteristics of all patients are reported in Table 5.

Multiple patients had more than one admission period, which is why admission characteristics are reported in a separate table. As models were trained and tested on different input signals, BG and extended, from different time periods, intra, pre and cross-admission, the amount of admission data that was available differed. Baseline admission characteristics were split to show characteristics based on data availability. Characteristics for BG are reported in Table 6. Extended characteristics are reported in Table 7.

Table 5 Baseline table for patient characteristics

Characteristic	Cases (n = 15)	Missing values
Female	10 (66.7%)	0
Age, Years, (mean [SD])	56.33 [15.9]	0
Type Diabetes		10
1	3 (20%)	
2	2 (13.3%)	
Unknown	10 (66.7%)	
Type sensor		0
Freestyle LibreLink	10 (66.7%)	
Medtronic	5 (33.3)	

HbA1c, mmol/mol, (mean [SD])	61.35 [8.95]	3
BMI, kg/m^2, (mean [SD])	28.58 [6.22]	3
BSA, m^2, (mean [SD])	1.89 [0.2]	8

*Table 6 Baseline table admission characteristics for patients with only CGM data. *One patient can have multiple admission periods and therefore multiple different admission specialisms.*

Characteristic	**Admission Cases (n = 19)**	**Pre-admission Cases (n = 27)**
Data duration, Days, (mean[SD])	2.5 [2.73]	7 [0]
Admission specialism (%) *		
Cardiology	1 (5.3%)	-
Surgery	6 (31.6%)	-
Gynaecology	3 (15.8%)	-
Internal Medicine	5 (26.3%)	-
Gastro-intestinal-liver	2 (10.5%)	-
Orthopaedics	1 (5.3%)	-
Urology	1 (5.3%)	-
TIR, %, (mean[SD])	62 [18]	67 [18]
TAR, %, (mean[SD])	34 [19]	30 [18]
TBR, %, (mean[SD])	4 [5]	3 [3]
Missing glucose values, %, (mean[SD])		
Total	2.56 [3.56]	5.49 [4.92]
Between 6:00-11:00	0.73 [2.58]	2.96 [4.22]
Between 11:00-16:00	1.35 [3.27]	3.20 [4.27]
Between 16:00-0:00	6.44 [11.44]	7.49 [7.65]
Between 0:00-6:00	1.03 [1.71]	6.91 [9.81]

Table 7 Baseline table admissions for patients with CGM, insulin, and carbohydrate data.

Characteristic	**Admission Cases (n = 4)**	**Pre-admission Cases (n = 5)**
Data duration, Days, (mean[SD])	1.26 [0.28]	7 [0]
Admission specialism (%)		
Surgery	1 (25%)	-
Gynaecology	1 (25%)	-
Orthopaedics	1 (25%)	-
Urology	1 (25%)	-
TIR, %, (mean[SD])	67 [20]	77 [9]
TAR, %, (mean[SD])	33 [20]	20 [8]
TBR, %, (mean[SD])	1 [1]	3 [3]

Mean TDD insulin, unit, (mean[SD])	79.34 [58.62]	55.60 [28.01]
Mean basal insulin/day, unit, (mean[SD])	32.87 [30.02]	16.13 [4.92]
Mean bolus insulin/day, unit, (mean[SD])	46.47 [29.35]	39.47 [24.45]
Mean carbohydrates/day, grams, (mean[SD])	129.56 [73.71]	163 [86.83]
Missing glucose values, %, (mean[SD])		
Total	0 [0]	3.08 [1.64]
Between 6:00-11:00	0 [0]	2.72 [3.31]
Between 11:00-16:00	0 [0]	5.19 [6.05]
Between 16:00-0:00	0 [0]	4.05 [2.63]
Between 0:00-6:00	0 [0]	0.32 [0.46]

2.4.2 Data selection

Results from the same models with different sensors as input were compared during the study setup. Both sensors with different sample frequencies show no significantly different MAE over all PH ($p = 0.22$). No resampling was performed on either of these CGM signals.

Data availability was variable between patients in the different groups of input signals. For example, in the patient group having intra-BG data, some patients had admission periods of 1 day ranging to 11 days. Although different amounts of data available in different groups make comparison less fair, it was chosen to include as much data as possible due to the small group of patients, especially during admission. Different data availability in days is reported in Table 8.

Table 8 Different data availability in the different input signal groups

Input signals	Intra-BG	Intra-extended	Pre-BG	Pre-extended	Cross-BG	Cross-extended
Available data, days	1-11	1-1.6	6-7	6-7	6-17	6-7.6

2.4.3 Adaptive Identification algorithm

While some patients had more consistent orders, others had a lot of variation. For the models trained on intra-BG data, for example, for all patients, the model order changed on average 15.26 (±4.3) times. The most frequently used model order within each subject accounted for an average of 41.69% (±13.99) of all instances. The distributions

of different p, d, and q orders of all patients in the intra-BG group are visualized in Figure 3.

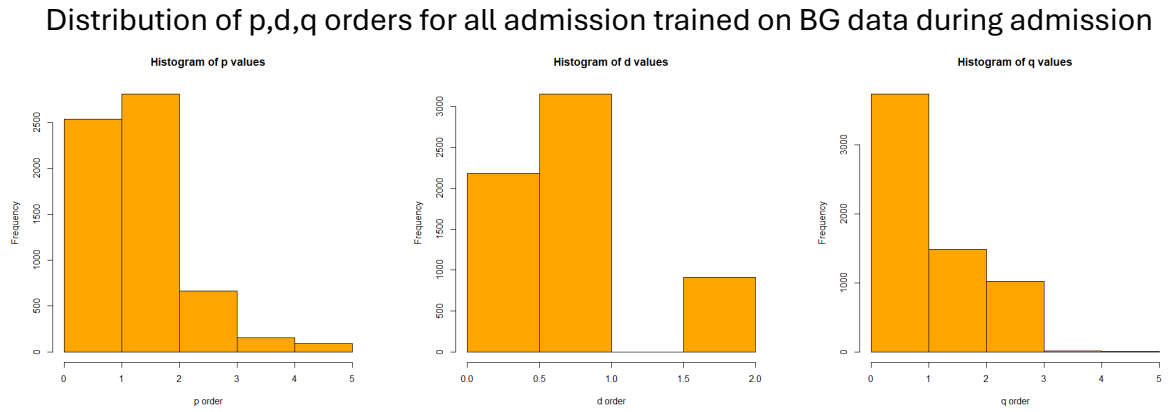


Figure 3 Distribution of p,d,q and q orders to display the variability of the orders and therefore the validation of the adaptive identification algorithm

An example of the results obtained by applying the adaptive identification can be seen in Figure 4. Here, results from the best-performing patient admission with the best-performing input signals, Cross-BG, are visualised. This patient admission is numbered patient admission 1.

2.4.4 Model residuals

Because of the high number of models and PH combinations, two key examples are reported, namely the residual diagnostics for the best and worst-performing patient admission-specific model for the input signals with the overall highest accuracy: Cross-BG. The best-performing patient admission-specific model for Cross-BG was patient admission 1. The worst-performing patient admission is patient admission 2. For the best-performing admission period, none of the residuals were approximately normally distributed with $p < 0.05$ for all PH. All residuals were not significantly autocorrelated with all $p < 0.05$. Histograms and residual vs. fitted plots for the best-performing admission are reported in Figure 5. For the worst-performing admission period, only for PH = 60 min. the residuals were approximately normally distributed with $p > 0.05$ for all PH. All residuals were not significantly autocorrelated with p-values < 0.05 for all PH. Histograms and residual vs. fitted plots for the worst-performing admission are reported in Figure 6.

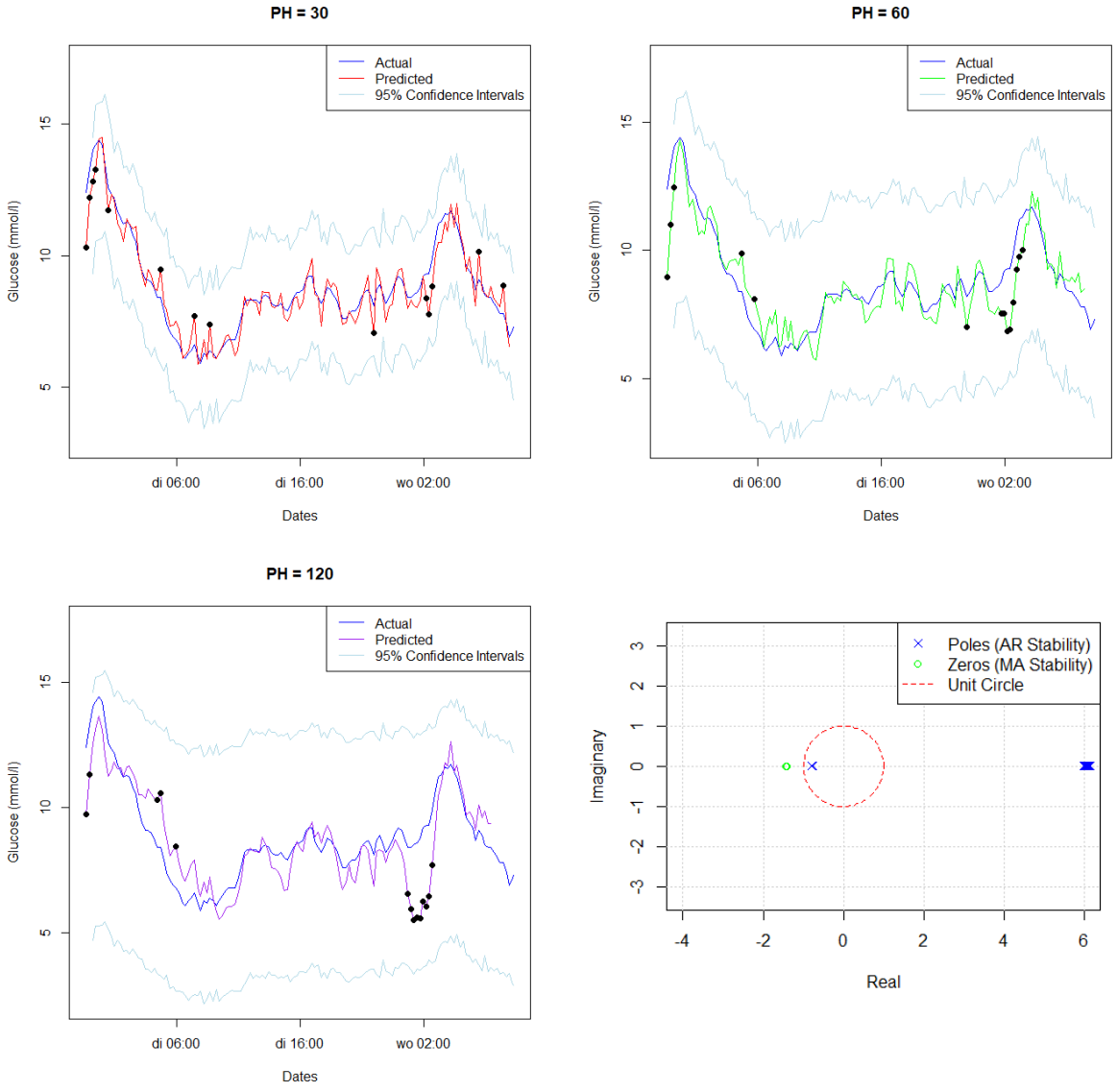


Figure 4 Predictions for a single patient for PH = 30 min, 60 min, and 120 min. The blue line is the reference signal, the red line is the predictions at PH = 30 min, the green line is the predictions with PH = 60 min, the purple line is the predictions at PH = 120 min, the light blue lines are the 95% confidence intervals, and black dots indicate the 10% biggest errors. On the right bottom the pole zero diagram for this patient is visualized with poles (blue cross) and zeros (green dots) for all identified models. The unit circle (red dotted line) is also visualized in the pole zero diagram

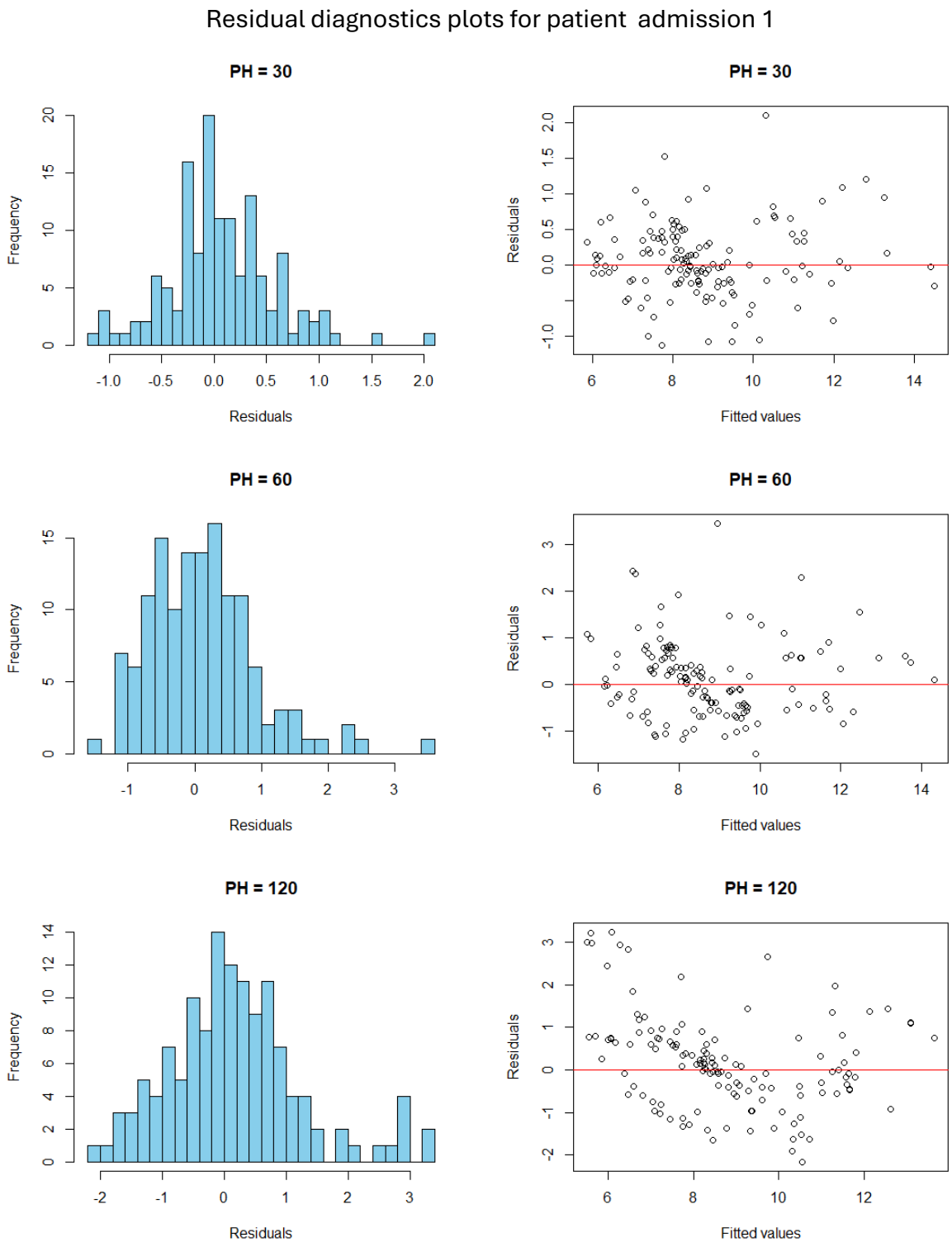


Figure 5 Histograms (left columns) and Fitted vs. Residual (right columns) plots for the best perfoming patient for the models trained on cross-BG. PH in minutes.

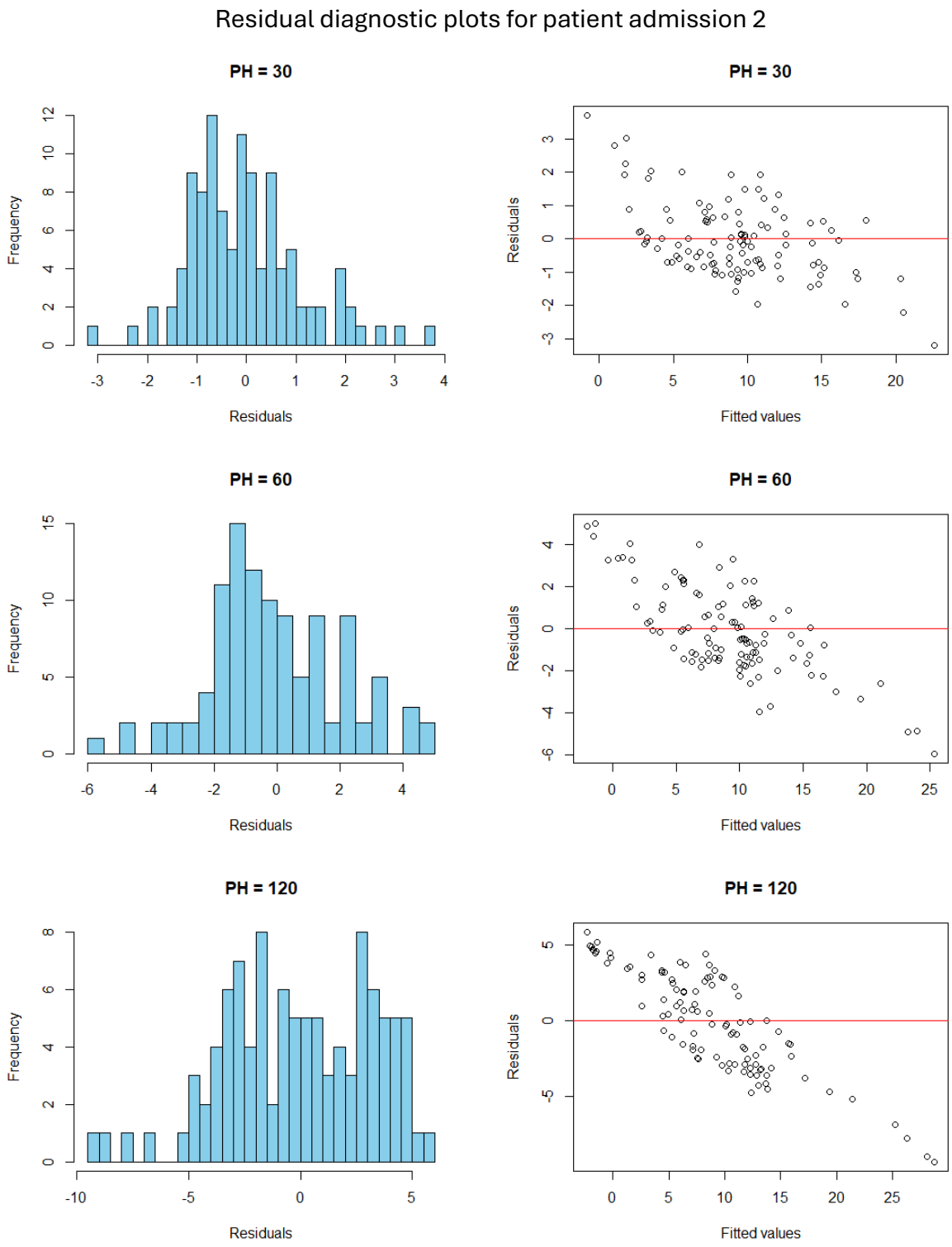


Figure 6 Histograms (left column) and Fitted vs. Residual (right column) plots for the worst performing patient for the models trained on BG pre-admission and during admission. PH in minutes.

2.4.5 Stability analysis

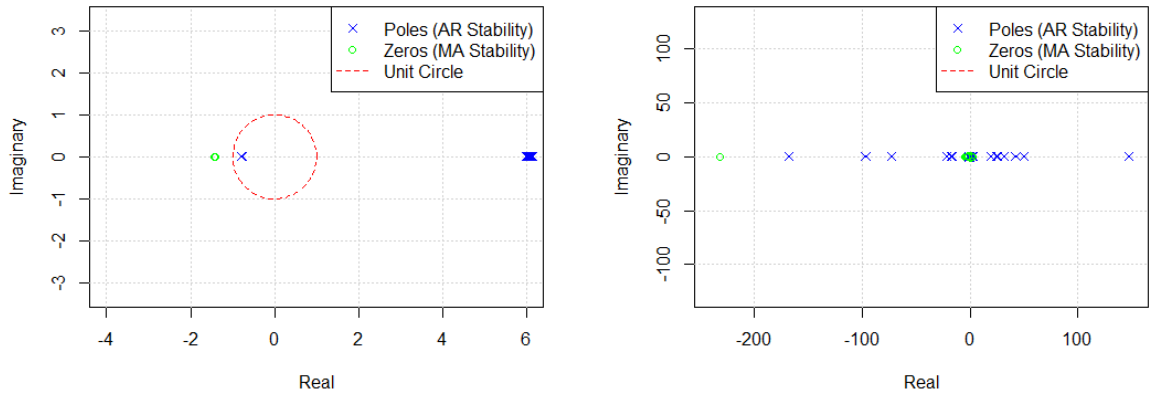
The asymptotic stability of the models was again tested for the best and worst-performing patient admissions with the cross-BG input signals as key examples. For patient admission 1, 139 models were made. This corresponds to the amount of data points available during admission. 138 of those models had the same (p,d,q) order, namely (2,0,1). The right side of Figure 7 shows that the best-performing admission period, patient admission 1, only had real poles and zeros. All poles and zeros from these models were analyzed. The number of real poles varied between 0 and 2. The maximum real pole value for this admission period was 6.14. The minimum real pole value was -0.79. The number of real zeros varied between 0 and 1. The maximum real zero value was -1.42. The minimum real zero value was -1.44. Because all models contained a pole outside of the unit circle, all models were unstable.

The left side of Figure 7 shows the worst-performing admission period, patient admission 2. For this admission period, 110 models were created, which correspond to the amount of data points available during the admission. (p,d,q) order varied in this patient admission, with the most frequent order being (3,1,5), which occurred 94 times. These models had real and complex poles and zeros. The number of real and complex poles varied between 0 and 2. The maximum real pole value for this admission period was 147.30, and the maximum complex pole magnitude was 1.09. The minimum real pole value was -168.28, and the minimum complex pole magnitude was 0.88. The number of real zeros varied between 0 and 2, and the number of complex zeros varied between 0 and 4. The maximum real zero value was 1.95, and the maximum magnitude of the complex zeros was 1.33. The minimum real zero value was -232.34, and the minimum complex zero magnitude was 1.02. Although some poles of the worst-performing admission model were within the unit circle, at least one pole was always outside, rendering the model unstable.

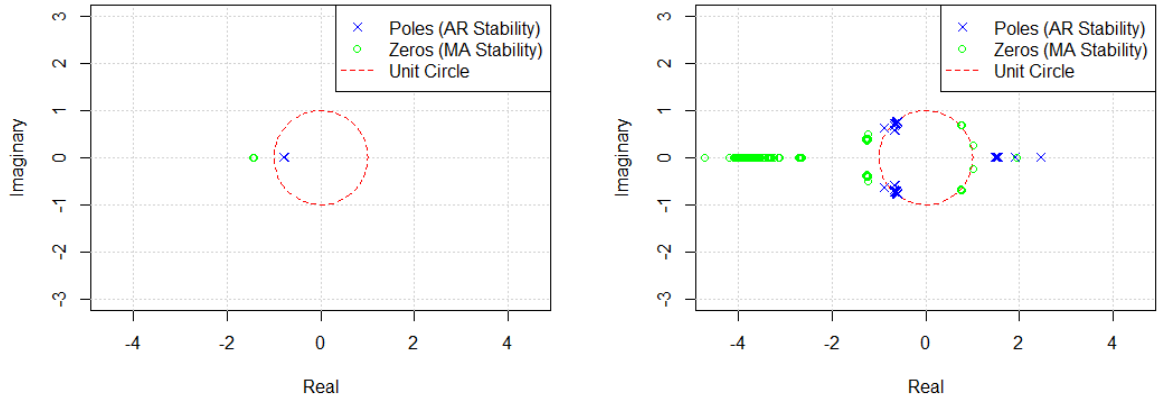
Poles and Zeros of all models over time

Patient admission 1 Patient admission 2

Zoomed out



Normal zoom



Zoomed in on unit circle

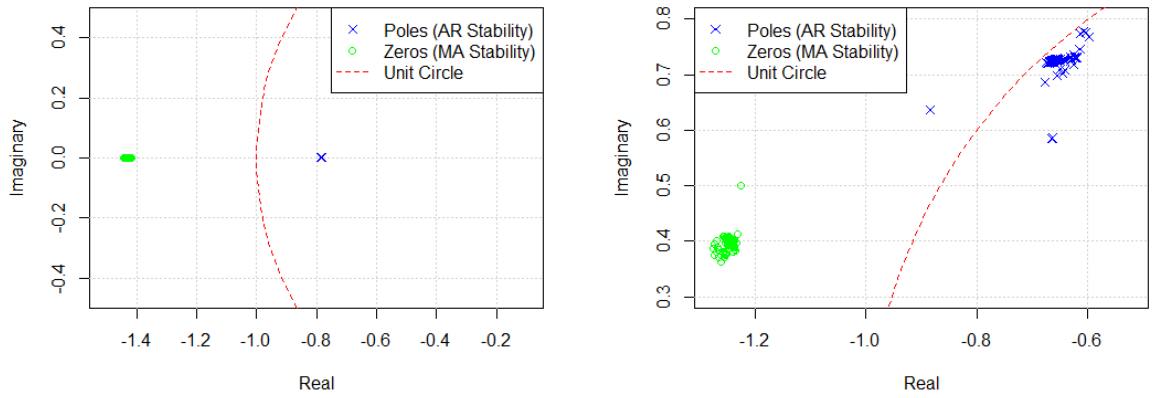


Figure 7 Poles and Zeros plot for best performing (left) and worst performing (right) patient admission with models trained on cross-BG data. On the top are figures zoomed out to contain all poles and zeros, including some unmeaningful outliers. In the middle is a medium zoom containing the largest part of all poles and zeros. On the bottom are two zoomed-in figures to see whether poles and zeros fall inside, on, or outside the unit circle. In the zoomed-in plots, only one side of the complex poles and zeros are shown as the other side contains complex conjugates.

2.4.6 Accuracy

The MAE provided by the adaptive identification algorithm, when trained on cross-BG data, showed the best average result of all PH. The MAE were, respectively, 0.65 (±0.21), 1.17 (±0.23), and 1.88 (±0.24) mmol/l for PH = 30 min, 60 min, and 120 min. For PH = 30 min., the result fell in the literature benchmark; for PH = 60 min. the result was below the literature benchmark, and for MAE, there was no benchmark for PH = 120. These results were acquired from seven admission instances. These seven instances resulted from the criteria set for data needed before and during admission. These seven admissions had at least 6 days of data before admission and at least 1 day during admission. All MAE results for all input signal combinations are reported in Table 9.

The RMSE provided by the adaptive identification algorithm showed the best average result of all PH on the cross-BG data. The RMSE was 0.87 (±0.27), 1.52 (±0.27), and 2.41 (±0.25) mmol/l for PH = 30 min, 60 min, 120 min respectively. For PH = 30 min. and PH = 60 min., the result fell within the literature benchmark, for PH = 120 min., the results were above the literature benchmark. RMSE results for all input signal combinations are reported in Table 10.

Table 9 MAE values from ARIMA(X) with adaptive identification algorithm for all PH, for all input signal combinations. The data that resulted in the best average result is highlighted.

MAE, mmol/l, (mean[SD])					
Input data	PH=30 min	PH=60 min	PH=120 min	Total	n
Intra-BG	0.67 [0.19]	1.21 [0.31]	2.10 [0.56]	1.26 [0.31]	19
Intra-extended	0.71 [0.07]	1.23 [0.16]	2.02 [0.34]	1.21 [0.15]	4
Pre-BG	0.78 [0.20]	1.42 [0.41]	2.22 [0.75]	1.42 [0.42]	27
Pre-extended	1.13 [0.52]	1.65 [0.62]	2.32 [0.83]	1.61 [0.59]	5
Cross-BG	**0.65 [0.21]**	**1.17 [0.23]**	**1.88 [0.24]**	**1.19 [0.18]**	**7**
Cross-extended	1.61 [0.72]	1.85 [0.38]	2.23 [0.47]	1.85 [0.33]	3

Table 10 RMSE values from ARIMA(X) with adaptive identification algorithm for all PH, for all input signal combinations. The data that resulted in the best average result is highlighted.

RMSE, mmol/l, (mean[SD])					
Input data	PH=30 min	PH=60 min	PH=120 min	Total	n
Intra-BG	0.92 [0.27]	1.62 [0.43]	2.76 [0.71]	1.39 [0.34]	19

Intra-extended	0.99 [0.07]	1.62 [0.19]	2.46 [0.42]	1.34 [0.18]	4
Pre-BG	1.08 [0.29]	1.91 [0.55]	2.93 [0.97]	1.55 [0.48]	27
Pre-extended	1.63 [0.74]	2.23 [0.79]	3.12 [1.00]	1.69 [0.59]	5
Cross-BG	**0.87 [0.27]**	**1.52 [0.27]**	**2.41 [0.25]**	**1.30 [0.19]**	**7**
Cross-extended	1.95 [0.86]	2.22 [0.49]	2.66 [0.58]	1.91 [0.28]	3

In almost all PHs, except PH = 120 min. for MAE, the combination of input signals had a significant effect on MAE and RMSE. Table 11 summarises the outcomes of ANOVA and post hoc analysis for MAE and RMSE.

Table 11 Outcomes of statistical analysis. Per metric, per PH, an F-value and p-value from ANOVA on the fixed effects (input signal combination) are given. Post hoc pairwise analysis is reported where the first reported input combination is x lower than the second reported input combination with a corresponding p-value.

Metric	PH (min.)	F-value	p-value (p<0.05)	Significant pairwise comparisons (average difference in mean metric value, mmol/l, corresponding p-value)
MAE	30	12.47	6.21e-8	Cross-BG vs. Cross-extended (-0.94, p<0.0001) Cross-BG vs. Pre-extended (-0.48, p<0.0001) Intra-BG vs. Pre-extended (-0.44, p=0.002) Intra-extended vs. Pre-extended (-0.43, p=0.04) Pre-BG vs. Pre-extended (-0.33, p=0.02)
	60	6.19	0.0002	Cross-BG vs. Cross-extended (-0.61, p=0.003) Cross-BG vs. Pre-BG (-0.28, p=0.007) Cross-BG vs. Pre-extended (-0.45, p=0.008)
	120	No significant effect		
RMSE	30	11.19	3.17e-7	Cross-BG vs. Cross-extended (-1.04, p<0.0001) Cross-BG vs. Pre-extended (-0.74, p<0.0001) Intra-BG vs. Pre-extended (-0.66, p=0.0002) Intra-extended vs. Pre-extended (-0.65, p=0.01) Pre-BG vs. Pre-extended (-0.5, p=0.005)
	60	6.17	0.0002	Cross-BG vs. Cross-extended (-0.6, p=0.04) Cross-BG vs. Pre-BG (-0.4, p=0.002) Cross-BG vs. Pre-extended (-0.66, p=0.002) Intra-BG vs. Pre-extended (-0.49, p=0.04)
	120	3.65	0.008	Cross-BG vs. Intra-BG (-0.5, p=0.03) Cross-BG vs. Pre-BG (-0.53, p=0.02)

2.4.7 Safety

Certain combinations between input signals and PH meet the safety criteria. Only the intra-extended data almost fulfilled the CEGA criterion for safety set in this study for all

PHs. These reported a total of 100%, 99.87%, and 99.42% in the A+B zones of the CEG for PH = 30 min, 60 min, 120 min respectively. The remaining predictions were in zone C. These results are visualized in Figure 8. All CEGA outcomes for different input signal combinations are reported in Table 12.

No input signal combination satisfies the set threshold for all PH when compared to the MAPE threshold of 10%. For PH = 30 min., the models trained and tested on intra-BG, pre-BG, cross-BG, and intra-extended data were compliant with the MAPE safety criterion. All MAPE values for different input signal combinations are reported in Table 13.

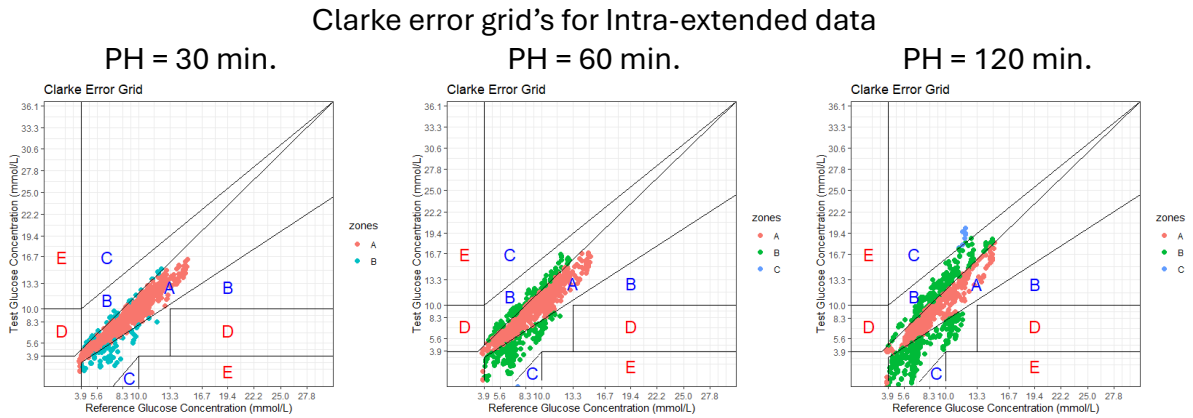


Figure 8 Clarke error grid results for intra-extended data for PH = {30 min, 60 min, 120 min}.

Table 12 CEGA results from ARIMA(X) with adaptive identification algorithm for all PH, for all input signal combinations

Predictions in CEG zones, %, (mean[SD])						
Input data	PH, min	A	B	C	D	E
Intra-BG	30	92.75 [5.19]	7.19 [5.15]	0	0.06 [0.22]	0
	60	78.45 [11.11]	20.56 [10.6]	0.43 [1.11]	0.54 [1.16]	0.02 [0.07]
	120	56.11 [12.45]	40.19 [11.19]	2.57 [2.6]	0.85 [1.92]	0.28 [0.71]
Intra-extended	**30**	**91.6 [2.93]**	**8.4 [2.93]**	**0**	**0**	**0**
	60	**75.62 [10.28]**	**24.27 [10.22]**	**0.11 [0.22]**	**0**	**0**
	120	**60.49 [11.05]**	**48.93 [10.22]**	**0.58 [0.87]**	**0**	**0**
Pre-BG	30	89.34 [5.42]	10.41 [5.16]	0.03 [0.08]	0.21 [0.45]	0.01 [0.03]
	60	71.44 [9.3]	27.37 [8.45]	0.57 [0.78]	0.58 [0.95]	0.03 [0.09]

	120	52.52 [9.7]	43.24 [7.62]	2.82 [3.1]	0.96 [1.15]	0.47 [0.66]
Pre-extended	30	78.89 [10.38]	19.56 [8.55]	0.26 [0.37]	1.25 [1.9]	0.04 [0.07]
	60	64.81 [11.21]	33.19 [9.06]	0.89 [1.22]	1.09 [1.63]	0.03 [0.07]
	120	51.44 [11.73]	43.7 [8.44]	3.63 [2.64]	0.85 [1.25]	0.38 [0.54]
Cross-BG	30	94.01 [5.18]	5.74 [4.82]	0.1 [0.5]	0.13 [0.61]	0.02 [0.12]
	60	80.8 [13.44]	18.1 [11.9]	0.1 [0.51]	0.93 [3.37]	0.07 [0.38]
	120	62.58 [16.19]	34.65 [14.34]	1.24 [2.63]	1.39 [3.25]	0.14 [0.45]
Cross-extended	30	75.38 [22.11]	22.72 [19.68]	0	1.9 [4.25]	0
	60	67.83 [18.64]	30.7 [17.27]	0	1.47 [3.29]	0
	120	56.34 [16.04]	42.57 [15.22]	0.26 [0.58]	0.84 [1.87]	0

Table 13 MAPE results from ARIMA(X) with adaptive identification algorithm for all PH, for all input signal combinations. The data that resulted in the best average result is highlighted.

MAPE, %, (mean[SD])			
Input data	PH = 30 min	PH = 60 min	PH = 120 min
Intra-BG	8.18[3.05]	15.08[5.51]	26.17[9.93]
Intra-extended	8.73[1.56]	15.41[3.11]	25.11[5.16]
Pre-BG	9.84 [2.82]	17.78 [4.99]	27.17 [7.80]
Pre-extended	14.57 [5.42]	21.13 [6.31]	29.16 [9.53]
Cross-BG	**7.66 [2.96]**	**14.03 [5.61]**	**22.98 [9.53]**
Cross-extended	13.36 [6.48]	18.50 [4.64]	26.26 [9.60]

For percentages in A and B zones in the CEG, the input signal combination had a significant effect with PH = 30 min. For MAPE, with PH = 30 min. and 60 min., input combination has a significant effect. Table 14 summarises the outcomes of ANOVA and post hoc analysis for MAE and RMSE.

Table 14 Outcomes of statistical analysis. Per metric, per PH, an F-value and p-value from ANOVA on the fixed effects (input signal combination) are given. Post hoc pairwise analysis is reported where the first reported input combination is x lower than the second reported input combination with a corresponding p-value.

Metric	PH (min.)	F-value	p-value (p<0.05)	Significant pairwise comparisons (average difference in mean metric value, %, corresponding p-value)
CEGA % in A+B zones	30	3.47	0.007	Cross-extended vs. Intra-BG (-1.96, p=0.02) Cross-extended vs. Pre-BG (-1.79 p=0.03)
	60	No significant effect		
	120	No significant effect		
MAPE	30	8.7	3.09e-6	Cross-BG vs. Cross-extended (-5.65, p=0.001) Cross-BG vs. Pre-BG (-2.13, p=0.04) Cross-BG vs. Pre-extended (-6.51, p=0.001) Intra-BG vs. Pre-extended (-6.33, p=0.0001) Intra-extended vs. Pre-extended (-6.56, p=0.005) Pre-BG vs. Pre-extended (-4.39, p=0.01)
	60	4.89	0.0009	Cross-BG vs. Pre-BG (-3.71, p=0.004) Cross-BG vs. Pre-extended (-6.03, p=0.01)
	120	No significant effect		

2.4.8 Usability

In terms of usability, the proposed ARIMA(X) model with an adaptive identification algorithm fell within the usability benchmark for all amount of training days that were analyzed. When applying different amounts of training data, no trend could be observed in the accuracy metrics averaged over all PH. Both the mean and standard deviation of MAE and RMSE seem to have no trend. The error values are reported in Table 15.

The ARIMA(X) model, utilizing an adaptive identification algorithm, can make predictions for a given PH when the training data length corresponds to the duration of that PH, provided that the model is trained solely on BG data. For example, the model only needs 30 minutes of data when making a first prediction for PH = 30 min.

The largest computational time was 0.001 seconds for one day of training data on the glucose, insulin, and carbohydrate data from pre-admission and admission. The computational time was approximately constant with an increasing number of training days.

Table 15 Accuracy metrics for different amounts of training data for all models trained Cross-BG and Cross-extended data. The MAE and RMSE are averaged over all PH to give a representative metric for all PH.

	Cross-BG, (n=7), mmol/l, (mean[SD])		**Cross-extended, (n=3), mmol/l, (mean[SD])**	
Amount of training data, Days	MAE	RMSE	MAE	RMSE
1	1.22[0.55]	1.59[0.66]	1.56[0.57]	1.95[0.61]
2	1.22[0.54]	1.58[0.66]	1.55 [0.62]	1.99[0.72]
3	1.21[0.53]	1.56[0.65]	1.62[0.56]	2.05[0.57]
4	1.21[0.54]	1.57[0.66]	1.53 [0.49]	1.94[0.51]
5	1.23[0.55]	1.58[0.67]	1.64[0.36]	2.04[0.36]
6	1.24[0.56]	1.59[0.68]	1.46[0.54]	1.86[0.61]

2.5 Discussion

This research was focused on assessing a clinically relevant BG prediction by developing different patient-specific ARIMA(X) models in an adaptive identification framework for persons with diabetes in non-ICU hospital wards. This was done by assessing the performance of the predictions of the developed models on accuracy, safety, and usability. Cross-BG data showed the most promising predictions that satisfied the different criteria and benchmarks of clinical relevance across various PHs. The satisfaction of the criteria and benchmarks for clinical relevance depended heavily on the different PHs. The sub-questions of this study are considered to provide detail to these results.

The first sub-question from this study was: which input signals result in the most clinically relevant BG prediction model for PH = 30 min, 60 min, and 120 min? Regarding MAE and RMSE, the combination of input signals had a significant effect in almost all PHs, except PH = 120 min. for MAE. Regarding the percentage in A+B zones in the CEG, the input signal combination had a significant effect on the outcome with PH = 30 min. For MAPE, with PH = 30 min. and 60 min., the input combination has a significant effect on the outcome. Cross-BG data showed the most promising combination predictions that satisfied the different criteria and benchmarks of clinical relevance across various PHs, all benchmarks and criteria are summarised in Table 16. Table 17 illustrates the compliance of the models trained on cross-BG data. In this table, compliance with the criteria and benchmarks is indicated with colors. Green stands for below a criterion or benchmark, indicating sufficient performance; orange indicates within the benchmark, suggesting comparable performance to existing literature; and red means above the criteria or benchmarks, indicating insufficient performance.

Table 16 Summary of literature benchmarks and criteria for clinical relevance assessment

Literature benchmark or criterion per metric	PH = 30 min.	PH = 60 min.	PH = 120 min.
Accuracy MAE	0.64 – 0.86 mmol/l	1.57 mmol/l	-
Accuracy RMSE	0.76 - 1.52 mmol/l	1.30 – 2.14 mmol/l	1.64 – 1.95 mmol/l
Safety MAPE	<10%	<10%	<10%
Safety CEGA	A+B zones	A+B zones	A+B zones
Usability training	Average RMSE across all PH inside: 1.23 – 1.87 mmol/l		
Usability computational	Considered usable based on expert opinion		

Table 17 Summary of compliance with set benchmarks or criteria for the models trained on cross-BG data. Green stands for below a criterion or benchmark, which indicates sufficient performance; orange indicates within the benchmark, suggesting comparable performance to existing literature; and red means above the criteria or benchmarks, indicating insufficient performance.

Criteria/Benchmark	PH = 30 min.	PH = 60 min.	PH = 120 min.
Accuracy MAE & RMSE	MAE: 0.65 mmol/l RMSE: 0.87 mmol/l	MAE: 1.17 mmol/l RMSE: 1.52 mmol/l	MAE: 1.88 mmol/l RMSE: 2.41 mmol/l
Safety CEGA	99.75% in A+B	98.9% in A+B	97.23% in A+B
Safety MAPE	7.66%	14.03%	22.98%
Usability training	MAE 1.22 mmol/l & RMSE 1.59 mmol/l with 1 day training data		
Usability computational	0.001 seconds of computational time		

Models with exclusively BG as an input signal perform better in general than those made with extended input signals, although not significant in all pairwise combinations. To obtain the input signal for carbohydrates, the insulin pump records carbohydrate intake through the manual entry of carbohydrate amounts by the user. It is hypothesized that a stressful hospital admission can lead to a more difficult scenario for the patients to handle their insulin pump settings and actively count their carbohydrates (Fortin et al., 2017; Pérez et al., 2020b). This would reduce the quality of the carbohydrate input signals, which could affect model performance when this is used as training data. From research by Sawaryn, it is known that carbohydrates can have a significant influence on BG prediction in patients-specific ARMAX models if perturbations are sufficiently high (Sawaryn, 2020). This effect is most pronounced in PH = 30 min. and 60 min, and for perturbations of 70 and 140 grams of carbohydrates. In addition, analysis of the baseline tables revealed that the differences in TIR, time above range (TAR), and time below range (TBR) metrics between pre-admission and during admission were notably smaller in the group with only BG data as opposed to the group with extended input signals. This difference in BG ranges before and during admission can cause more randomness in BG

data patterns for persons with cross-extended data, which can lead to more erroneous predictions(Hyndman, 2007).

For pre-extended, cross-extended, and PH = 60 min. and 120 min. the developed models do not predict BG in a clinically relevant way. This is largely due to the constraint that the MAPE must be smaller than 10%. From the performed literature research, no study was found that had a similar metric. The metric is assumed to be important in this research as a MAPE of more than 10% can lead to a significant increase in hyper- and hypoglycemic events (Kovatchev et al., 2015). The occurrence of these events is exactly what this study tries to prevent in an in-hospital scenario.

The second sub-question from this study was: what amount of training data results in the most clinically relevant BG prediction model for PH = 30 min, 60 min, and 120 min? Cross-BG and Cross-extended data showed that the accuracy of the predictions averaged across all PHs stays approximately constant with one to six days of training data, and the accuracy metrics fell within the literature benchmark. This means that for both combinations of input signals, one day of training results in sufficient accuracy. Although the averaged accuracy metrics and literature benchmarks provide a convenient summary of usability performance across PHs, this approach may obscure variations specific to individual PH. The interpretation of these results should account for this potential loss of information inherent to averaging these metrics and the benchmark. This data is useful to determine the usability of the proposed model. In a hospital admission scenario, this means that one day of training data is sufficient to start making predictions. This implies that the model can be used for short and long admission stays. In addition, if one day of BG data is available before admission and the model is trained using that data, the predictions can start right away during admission.

To reduce the calculation time of each iteration, a sliding training data window can be imposed. This sliding window would obviate the need for repetitive model training with all available data(Yang, 2019). As MAE and RMSE are constant for one to six training days in this study, a window of one day can be imposed. This follows the method for window selection from Yang et al. (Yang, 2019). The window of one day approximately corresponds with the findings by Yang et al. as they impose a window length of 20 hours.

The last sub-question focuses on what the added value is of a patient-specific adaptive identification approach for BG prediction with ARIMA(X) models. Due to high patient inter-variability in BG trends in persons with diabetes, a patient-specific modeling approach was chosen in this study (Laguna et al., 2014). Mohebbi et al. used a population-based long-short-term network to predict BG with the PH = 30 min. and 60 min. (Mohebbi et al., 2020). Mohebbi et al. assessed the predictions based on MAE and RMSE and compared them to patient-specific ARIMA models. When the population-based models were tested against patient-specific ARIMA models on the same dataset, the population-based model outperformed the patient-specific ARIMA model with 7

days of training data. With little training data, the patient-specific ARIMA model outperforms the population-based model. The models in this study outperform the population-based models from Mohebbi et al. However, the data context is different. In this study, the limited data availability makes the patient-specific ARIMA approach favorable. One benefit of population-based models is that they do not require training for each individual patient. Consequently, models can be used for patients who have not previously had a CGM sensor before admission without needing training on their specific data.

In individuals with diabetes, not only is there variation between patients, but also within the same patient. To respect the intra-patient variability, resulting in time-varying BG dynamics and sometimes non-stationarity, an adaptive identification algorithm to adaptively determine (p,d,q) order and parameters of the ARIMA(X) models was implemented (Boiroux et al., 2017; Yang, 2019). This method is justified by the order (p,d,q), which often changes over time in this study. However, it should be noted that this approach can lead to overfitting. This overfitting can lead to unstable models in scenarios where sudden large changes happen in the BG signal. If this large change is an outlier or noisy data, this can lead to a failure of the model to grasp the real underlying dynamics, which in turn leads to inaccurate and possibly unsafe predictions. In this study, the best and worst-performing patient admission periods showed only unstable models, which could be a sign of this overfitting. To counteract this overfitting, the AIC criterion is applied to all newly formed models.

The instability of the models can influence BG predictions with a larger PH. PH = 30 min. outperforms PH = 60 min. and 120 min. in this study. This could be because of the recursive nature of forecasting with ARIMA(X) models. Prediction errors are compounded by increasing PH, which causes less reliable predictions (Selim et al., 2020). Instability does not have to be a problem. A BG signal is bounded in a certain physiological range due to the treatment of high and low BG values, which decreases the exploding effect an unstable model can have on the predictions. In addition, the short-term PH with a new prediction at every new time point in this study ensures that the effect of error propagation is limited. In a future study, ARIMA(X) models without this adaptive algorithm can be compared with this study's approach to see whether the effect of possible overfitting and instability has a worse effect on clinical relevance in prediction than disregarding intra-patient variability. Furthermore, other techniques to prevent overfitting in outlier scenarios can be researched.

The ARIMA(X) models with adaptive identification were developed to have a positive implication on tighter glucose regulation for inpatients without increasing hypoglycemia risks. PH = 30 min. shows promising results with the currently developed models, which can have clinically relevant implications. If a healthcare professional or a patient can be alarmed 30 minutes before a hyper- or hypoglycemic event, it enables them to take

proactive action to avert critical events (Prendin et al., 2021). This, in combination with the small amount of data needed for accurate and safe prediction, makes the proposed algorithm with PH = 30 min. useful for implementation in the clinic.

This study fills a literature gap. The adaptive identification algorithm was tested on patient data from a hospital admission environment. Besides this, a broader set of performance metrics on accuracy, safety, and usability and a choice of PH until 120 make this study tailored toward a practical solution to the challenges discussed. MAPE as a metric for safety sets this study apart from others as it assesses the error more strictly than CEGA. CEGA, however, already compares the possible final application to the ISO 15197 standard of >99% of possible measurements in the A+B zones of the CEG (Freckmann et al., 2019). As most other literature only discusses accuracy metrics, mainly RMSE, this study can be compared in that aspect to the literature. The RMSE of this study outperforms all studies except for the study by Sawaryn. Sawaryn also outperforms this study on safety metrics as CEGA shows higher percentages in A+B zones. However, as Yang et al. showed when comparing their results from T2D to T1D, T2D has a lower error in their predictions. In this study, T1D and T2D patients are mixed, which could be a reason for this result.

There are limitations to this study. Only a small sample of admission data was available. From the small sample, little was known about the patient characteristics or the course of the admission. The medical admission records from persons included in this study were examined. There was insufficient data to draw conclusions about significant factors during their admission that might affect model behavior. For different sensors, BG samples were taken in different frequencies, namely every 5 (Medtronic) or 15 minutes (FreeStyle Libre Link). Not only does that make the data pattern different, but it also implies different treatment. The Medtronic is an AID system, while the FreeStyle Libre Link is a standalone CGM system. A quick analysis of accuracy metrics for different sensor sample frequencies showed no significant differences, which is why it was chosen to apply the same method to both sensors and patient groups. However, future research could explore resampling the sensor signal measured every 15 minutes to match the signal measured every 5 minutes. Different interpolation methods could be applied to see if the predictions made with the resampled signal outperform the original predictions. During the included admission periods, there was no corticosteroid medication or recorded enteral or parenteral nutrition, while these factors influence the BG inpatient BG regulation (Pérez et al., 2020b). Before applying the models in admission cases involving corticosteroid medication, testing should be conducted first.

To improve the methods used, a few recommendations are proposed. Beginning with data collection, a more comprehensive understanding of the algorithm's clinically relevant performance can be achieved as additional data is gathered during admission. This way, something can be said about what events or factors influence prediction

during admission. It is recommended to track insulin infusion and nutritional intake to be certain of the sensitivity of the models for these factors. Another option to gain more insight into model behavior could be simulated data. This allows for more elaborate testing on what data characteristics benefit the model or impose risks. A more robust method for training ARIMA(X) models might be beneficial to prevent overfitting of the models. A training method proposed by Sawaryn could be applied in which the training set is cut into three with minimum periods of 1 day. For each part, the model with the lowest mean AIC was determined, and that model was considered to have the optimal combination of parameters. The simplest model remains valid if its AIC is no more than two units away from the average AIC. So, the model within that AIC range is chosen with the least number of parameters. This method reduces the risk of overfitting and ensures less data to fine-tune the model (Sawaryn, 2020). With more robust data acquisition and methods, models are hypothesized to become more stable, and therefore, predictions for PH = 60 min and 120 min are expected to become more clinically relevant. Even though PH = 30 min. already has a promising implication on inpatient care, PH = 60 min and 120 min would lead to even more significant advantages. Bolus calculators in AID can recommend a bolus amount based on the patient's carbohydrate intake and insulin sensitivity (Boiroux et al., 2017). The BG prediction algorithm from this study could be used as a state estimator in model predictive control (MPC). This means the model should also simulate the sensitivity of insulin and carbohydrates. For this implementation, the length of the PH should match the length of the effect of these effector variables, which for fast-acting insulin is at least 120 minutes (Crecil Dias et al., 2020)(Summary of Product Characteristics: NovoRapid, n.d.). Sawaryn showed that for PH = 30 min. and PH = 60 min., carbohydrates can significantly affect ARMAX predictions (Sawaryn, 2020). For the model's prediction sensitivity to insulin perturbations, similar research to Sawaryn can be performed, with extra safety concerns or in a retrospective manner. A simpler approach could already be initiated by applying the insulin dosing algorithm for inpatients by Pérez et al. with the current ARIMA model for PH = 30 min. to give proactive bolus insulin (Pérez et al., 2020b).

These recommendations pose multiple opportunities for future research. The adaptive identification algorithm could be studied more elaborately on purposefully collected and simulated patient data. The same study could also analyze the models' sensitivity to insulin and carbohydrate inputs, akin to the analysis performed by Sawaryn (Sawaryn, 2020). For this study to have direct implications, a study could be performed on the use of the model as a decision-support system either only as a 30-minute-ahead alarm or in combination with the bolus algorithm of Pérez et al. under strict expert supervision.

2.6 Conclusion

This research was focused on assessing a clinically relevant BG prediction by developing different patient-specific ARIMA(X) models in an adaptive identification framework for persons with diabetes in non-ICU hospital wards. This was done by assessing the performance of the predictions of the developed models on accuracy, safety, and usability. Cross-BG data showed the most promising combination predictions that satisfied the different criteria and benchmarks of clinical relevance across various PHs. The satisfaction of the criteria and benchmarks for clinical relevance depended heavily on the different PHs. One day of training results in sufficient accuracy. Due to high intra- and inter-patient variability, the patient-specific ARIMA(X) models with adaptive identification were justified. Only PH = 30 min. shows potential with the current algorithm, offering clinically relevant implications for managing hyper- and hypoglycemic events. Recommendations for improving methods are made to enable further research and implementation of the decision-support tool in the hospital. A decision-support tool could aid in tighter BG control in the hospital.

3. Chapter 2: Challenges for doctors in glucose regulation in hospitalized patients: a qualitative insight from medical doctors working in the department of internal medicine

3.1 Abstract

Introduction: To create a decision-support system that improves glucose regulation and ensures effective hospital implementation, understanding end users' perspectives is crucial. A small requirement engineering study to explore requirements with doctors in Dutch hospitals gathers insights into doctors' challenges with current BG practices.

Methods: This qualitative research involved semi-structured interviews with current or former internal medicine doctors. The interviews examined obstacles and methods in BG regulation. Two researchers analyzed the data: the first used inductive thematic analysis to assign descriptive open codes to key text fragments. This was followed by axial coding for organization. A second researcher verified the codes, ensuring consistency.

Results: This study contained a group of 7 doctors (42.9 % female) from multiple centers. Most of them were junior internists (57.1%). Key obstacles include insufficient patient data, such as nutritional status and home diabetes medication, and lack of knowledge among other specialties and care staff. Add-on insulin schemes delay BG regulation. Corticosteroid use complicates BG management. Interviews revealed similarities in necessary data for BG regulation policy: current/past BG values, home diabetes medication, corticosteroids, nutritional status, and relevant past history. Feedback on current practices varied; some found them effective, while others faced challenges.

Discussion: Obstacles, current methods, and perceived solutions suggest implementing a digital tool for data visualization and decision-support. It's important to consider these expert insights in the tool's development. Additionally, existing literature supports the need for such a tool. A strong grasp of the issues perceived by the target audience can result in more effective tool development and adaptation.

3.2 Introduction

To create a decision-support system for improved glucose regulation and effectively implement it in a hospital setting, understanding end users' perspectives on the current challenges is crucial. A literature search showed one article by Rousseau et al., performing an interprofessional qualitative study of barriers and potential solutions for the safe use of insulin in the hospital setting (Rousseau et al., 2014). The study identified

several challenges and potential solutions relevant to assessing requirements for a decision-support system in BG regulation among healthcare workers. In the study by Rousseau et al., a lack of documented insulin doses in electronic health records (EHR), lack of data points in BG measurement, and unavailability of nutritional data in EHR were a few of the perceived problems. A possible solution was the availability of insulin decision-support systems. This can be seen as a first indication of a likely adaptation of the decision-support system outlined in this report for healthcare professionals.

Since the publication of the study by Rousseau et al., there have been advancements in EHR and diabetes technology. Besides this, a single study done in a single center is not enough requirement analysis for the start of development and implementation of the decision-support tool. To address this gap, a small requirement engineering study to explore requirements was conducted with doctors in Dutch hospitals to understand their views on current obstacles in BG regulation practices. Exploration of requirements is part of requirement engineering, which is the discipline that covers creating and reporting requirements. This is essential for the development and implementation of software applications (Boulanger, 2017). For the purpose of exploring requirements for the decision-support system, this study focuses on the obstacles and current methodsin BG regulation to achieve the following objectives:

Explore the key obstacles faced by doctors and understand their methods of daily glucose regulation in inpatients.

3.3 Methods

To achieve the objective, qualitative research was conducted through semi-structured interviews. This structure was selected to identify genuine pain points in current practices and to gain a deeper understanding of these pain points and practices.

The study participants were medical doctors currently or previously employed in the internal medicine department, thus responsible for overseeing BG regulation in the hospital. The study was performed in the context of a study program in business validation. This program lasted for a short period of time, and several topics needed to be covered. The number of participants depended on how many volunteers could be found in a period of three days. This resulted in a total of 7 participants from the researchers' own network. The research subjects' composition was chosen to have a mix of people currently responsible for the BG regulation and those who supervise them. Senior medical interns, doctors not in training, junior doctors, and internists were interviewed.

The interviews took place from the 19th of July 2024 until the 21st of July 2024, either in an online meeting or by telephone. Prior to participation, participants received short information on the rationale behind the study but not on the concrete idea of the decision-support tool in order not to bias them. Answers were recorded by means of a

single researcher typing out the answers while simultaneously conducting the interviews. A single interview took about 30 minutes. Following the interview, participants were given the option to stay informed about the decision-support tool's development.

The interview consisted of questions on the main theme: obstacles and methods in BG regulation. At the start of each interview, each participant was introduced to the theme and context of the interview. This was followed by a very broad question about perceived obstacles. These questions were usually followed by some follow-up questions or repetition of the same question until a participant said to have named all perceived obstacles. This question was followed by two questions about the current methods of working on BG regulation in the participants' hospitals. The last topic concerned the possible hospitals' future goals for BG regulation. The interview ended with a question regarding any overlooked themes or further questions and comments from the participant, ending with a note of appreciation for their time. The full interview questions are reported in the Appendix.

Two researchers analyzed the interview data. The first researcher analyzed the interview using inductive thematic analysis using Microsoft Word and the comment function. Text fragments providing key insights on the research objective were assigned descriptive, open codes. These open codes underwent analysis and organization through axial coding, which involved comparing and merging them into broader codes that serve as themes and sub-themes. A second researcher revised the assigned codes, and no discrepancies were found.

3.4 Results

This study contained a group of 7 doctors (42.9 % female). Most of them were junior internists (57.1%). Hospitals of employment were diverse and can be seen along the rest of the characteristics in Table 18.

Table 18 Baseline table for participant characteristics

Characteristic	**Cases (n = 7)**
Female	3 (42.9%)
Role	
Internist – Endocrinologist	1 (14.3%)
Junior internists	4 (57.1%)
Doctor not in training internal medicine	1 (14.3%)
Semi-arts Internal medicine	1 (14.3%)
Hospital	
CWZ, Nijmegen	1 (14.3%)
Erasmus MC, Rotterdam	1 (14.3%)
Maxima MC, Veldhoven	1 (14.3%)
MST, Enschede	1 (14.3%)

Reinier de Graaf gasthuis, Delft	1 (14.3%)
ZGT Almelo	2 (28.6%)

3.4.1 Obstacles

The perceived obstacles were addressed. All participants agreed on factors causing deregulation in hospitals, especially on corticosteroid medication. All participants also noted the missing of specific data crucial for obtaining a clear view of the patient. Data on nutritional status and medication at home were most often missed. There was a lot of variety when it came to the perception of how difficult it was to handle deregulated patients. Different perceptions were as follows: *"This problem is very serious. It is always turned into our departments' problem. I seriously doubted whether I wanted to stay at internal medicine due to all the hassle around people with diabetes." [Participant 3].* On the other hand, a participant stated: *"It does not result in any problem around workload, but extra digitalization of data would be nice." [Participant 2].* The identified themes and sub-themes, with corresponding definitions and summarized results, are reported in Table 19.

Table 19 Themes and sub-themes with their definition and results from interviews for the topic of obstacles in BG regulation.

Themes and sub-themes	**Definition and/or result**
Causes of deregulation	
General causes	High BG at start of admission Nerves Different nutritional patterns
Medication	Especially corticosteroid medication (examples given were especially Prednisone and Dexamethasone)
Less chance of deregulation	Only oral medication (specifically Metformin)
Problems	
Time management	The impact of handling glucose regulation on time management in general. Most often glucose regulation was indicated as a frustrating problem due to the number of people with diabetes, the irregularity in planning visits for BG regulation, lack of knowledge in other specialists and care staff, and lack of staff.
Knowledge	The lack of knowledge when seeing a patient on the context of that specific patient.

	The lack of knowledge of other specialism on glucose regulation
Data	The problems that data can cause when regulating BG. This often comes down to a lack of data or difficulties finding the right data. Specific examples were the lack of data on nutritional patterns and modalities and (home) medication. On the difficulties finding data, one hospital in specific talked about data on paper leading to faulty regulation and the fact that data was inconveniently transferred from acute wards to long-term wards.
Hospital and EHR policy	Conservativeness in hospital and EHR policies make it hard to improve BG regulation.
Insulin treatment	An add-on insulin scheme[2] was indicated as falling behind instead of pro-active treatment leading to insufficient BG regulation.
Results of current glucose regulation algorithms	
Results	Perceptions on success rate of current algorithm on keeping patients within set BG ranges. Although perception was divided on this topic, the general perceptions was that on the department of internal medicine, current algorithms worked quite well. On other departments however this was not the case.
Patient groups	
Patient groups	Groups of patients susceptible of deregulated BG Patient with enteral/parenteral nutrition, corticosteroid medication, ketoacidosis, and therapy infidelity were seen as susceptible.

[2] An add-on insulin scheme is a sliding scale to determine the amount of insulin that can be added on to the regular diabetes medication at any moment when BG is measured. A common add-on insulin scheme is the 2-4-6 scheme. If this scheme applies, at the moment of BG measurement, and the BG values is >10mmol/l a patient gets 2 extra units of insulin, >15 mmol/l a patient gets 4 units of insulin, and >20 mmol/l a patient gets 6 units of insulin. The amount of insulin can be determined by estimated insulin sensitivity of a patient.

	Patients with AID or during pregnancy were identified as least susceptible.
Reliability literature	
Reliability literature	Perceived reliability of observational study results relating hyperglycemia during admission to adverse events for admitted patients. Everybody perceived these study results the same, as a chicken-and-egg-story. They cannot say for sure whether high BG led to adverse events or vice-versa. This had the implication that participants were not convinced to change their current methods because of the adverse events described by literature.
Relevance of the problem	
Relevance of the problem	The relevance of the problem was tested by participants perception on how much the BG regulation concerns them. No unambiguous perception was given. While some found it extremely concerning, others did not seem to perceive any problem with BG regulation as currently performed.

3.4.2 Methods of working

This section of the findings examines the existing working methods of the participants. The identified themes, with corresponding definitions and summarized results, are reported in Table 20.

Table 20 Themes and their definition and results from interviews on the topic of methods of working.

Themes	Definition and result
General	General comment on methods of working. Participants made mentions of protocols being used: perioperative, pregnancy, IV insulin pump, diabetes in hospital Add-on insulin schemes were mentioned often to regulate patients and to deduct

	the amount of long-acting insulin if needed.
Target value glucose	The different perceived BG target values for different hospitals. There were discrepancies between the different hospitals when it came to BG target values. Most of the hospitals had a range between 4-10 mmol/l, others between 4-12 mmol/l, and for some made a discrepancy between adult patient and elderly patients where elderly patients were kept between 4-15 mmol/l.
Responsible doctors	Responsible doctors for BG regulation in the hospitals. All participants mentioned the junior doctors or doctors-not-in-training under supervision of internists as responsible.
Measurement methods	Methods of BG measurements in hospital. All hospitals used either finger stick measurements or subcutaneous glucose sensors. Finger stick measurements were seen as old-fashioned and straining.
Necessary data	Data necessary to make policy decisions on hospitalized patients. The corresponding data named by all participants were: Current and past glucose values, (home) diabetes medication, corticosteroid medication, nutritional status, and admission reason. Other data named was: HbA1c lab value, can the patient self-regulate, medical past history, insulin-carbohydrate ratio. As to the questions whether participants preferred either CGM or finger stick BG results, everybody agreed that finger stick 4 times a day was enough data, however CGM is preferred when available.

3.4.3 Future goals

The final part of the results addressed future goals for BG regulation in the hospital. All solutions were proposed spontaneously by participants when talking about future goals. The identified themes, with corresponding definitions and summarized results, are reported in Table 21.

Table 21 Themes and their definition and results from interviews on the topic of future goals regarding BG regulation.

Themes	Definition and result
Solutions	Possible solutions for the perceived problems with BG regulations in the hospital. Most proposed solutions pointed in the direction of a digital overview of data needed for policy decisions. Other directions were detecting patients with high risk of deregulation in time and anticipate on this instead of add-on insulin schemes. Single suggestions were made to give diabetes nurses a bigger role in BG regulation in hospitalized patients, improve knowledge in care staff, and to have less invasive BG sensing incorporated.
Methods of working	Preferred extra methods of working not in the current methods. A comment was made on after-care for patients leaving the hospital.

3.5 Discussion

This study qualitatively explored the key obstacles faced by doctors and in addition, focused on understanding their current methods for achieving tight glucose regulation in inpatients.

Interviews reveal several key obstacles. All 7 participants mentioned one or more key obstacles. More obstacles were mentioned, particularly by those responsible for managing diabetes inpatients directly, that is, not as supervisors. The initial major challenge is the insufficient data to understand the patients' context, which was mentioned 9 times by 5 participants. Specifically, missing information includes their nutritional status and (home) diabetes medication, which were both mentioned by 5 participants. Additionally, a lack of knowledge and initiative to regulate BG among other

specialties and care staff has been highlighted as a significant barrier, mentioned by 3 participants. Furthermore, add-on insulin schemes are seen as important challenges since they lead to delays in BG regulation, which was mentioned by 2 participants. Finally, corticosteroid medication has been identified as a critical obstacle to managing BG levels, as mentioned by 5 participants.

Interviews provided extensive insights into the hospital's current methods for regulating BG. The necessary data for determining BG policy show quite a few similarities. Current and past BG values (home) diabetes medication were mentioned most often by 6 participants. Corticosteroid medication and nutritional status followed as the second most mentioned by 3 participants. Two participants identified past history and admission reasons as key data. Participants provided diverse feedback regarding the effectiveness of their current practices. Although two participants saw no issues, five found the existing BG regulation practices quite challenging or ineffective.

The varied responses to perceptions of the effectiveness of current practice could be explained by participants' roles. Supervisors or experienced doctors perceived fewer problems, while less experienced doctors perceived more problems. Even though essential information for policy development largely matched, it's challenging to determine if participants provided a comprehensive perspective. It is notable that participants did not see CGM as essential for effective BG regulation in the hospital. CGM is expected to improve glucose regulation in the hospital in the future (Zelada et al., 2023). The availability of CGM data was viewed as a superfluous luxury, as opposed to only four available finger stick BG measurements per day. To perfect the BG regulation, participants indicated that CGM data comes in handy when looking at home regulation and comparing it to in-hospital regulation or using it to estimate insulin sensitivity. Some participants questioned the non-proactive nature of add-on insulin schemes. However, they do not see a way to get rid of them as they are used as a titration tool for insulin amounts in their current practices.

A key perceived challenge is the insufficient data to make inpatient policies, particularly regarding nutritional information and home insulin protocols. This, combined with the obstacle of lack of knowledge in other specialties and among care personnel, highlights the absence of streamlined multidisciplinary protocols and centralized data, as the literature also points out this obstacle (Rousseau et al., 2014). Additionally, the frequent issue of corticosteroid medication causing deregulation in patients highlights the need for improved decision-support for doctors treating these patients. This is important to consider when developing a decision-support tool.

Since most perceived solution directions suggest implementing a digital tool for data visualization and decision-support, it's important to consider these expert insights in the tool's development. Additionally, existing literature supports the need for such a tool. A

strong grasp of the issues perceived by the target audience can result in more effective tool development and adaptation.

This short study has both strengths and limitations. Due to the limited time for data collection, only a few participants have been interviewed, and data saturation was not yet reached. However, this is no problem due to the explorative nature of this study. To further explore the findings from this study, more interviews can be held to reach data saturation. Three extra interviews will be conducted. If data saturation is not yet reached after this, another three interviews will be conducted, and so on (Francis et al., 2010; Malterud et al., 2016). The interview questions stem from a course in problem validation rather than from research literature. The answers to the questions were recorded solely through note-taking and were not transcribed from audio files. Despite a limited number of participants, a key strength of this study was the diversity of participants' hospitals.

The limitations highlight the potential for improvement in future research. A larger-scale study could incorporate additional, literature-supported questions to provide a broader understanding of perceived obstacles. Furthermore, the results from this study can guide the early development of a decision-support tool designed to assist doctors in effectively managing tight BG regulation for inpatients. The emphasis should be on integrating all required data, enhancing the multidisciplinary approach to BG regulation, and aiding doctors in preventing adverse events among complex patients.

3.6 Conclusion

This study identifies obstacles and insights into BG regulation practices. A major challenge is insufficient patient context data, specifically on nutritional status and diabetes medications. Additionally, a lack of awareness among specialties is a significant barrier, as are add-on insulin schemes causing delays in BG regulation. Corticosteroid medication also critically hinders BG management. Necessary data for BG policy reveal similarities, including current and past BG values, medications, nutritional status, and admission reasons. Participants provided varied feedback on the effectiveness of BG regulation practices; some saw no issues, while others faced significant challenges. Early results can inform the development of a decision-support tool to help doctors manage tight blood glucose levels for inpatients by integrating all necessary data and enhancing multidisciplinary approaches to prevent adverse events in complex patients.

4. Chapter 3: The patient's story: two patient personas

4.1 Abstract

Introduction: Patient stories and perspectives can help to understand this stakeholders' characteristics better, accommodate a system to different patient needs, increase empathy among users, and make the results of this report reality-centered and tangible. Personas can be utilized to obtain the patient's perspective.

Methods: Data was collected between the 8th of January and the 9th of December 2024 during clinical rounds for patients who needed extra consultation regarding their BG regulation during admission to the ZGT Almelo. Admission had to be relevant to the topic of the decision-support system and last more than 2 days to be included in the personas. Conversations with patients were held, and EHR was checked to obtain data. Data was gathered on the reason for admission, admission department, relevant past history, medication, current history information, policy information, and re-admission. From this data, fictional personas were developed by a single researcher to guarantee anonymity.

Results: Two personas were created to help understand patient characteristics and create empathy for patients' situations. From these personas, challenges were highlighted in which the decision support system could aid their glucose regulation during admission.

Discussion: The personas illustrate that it is difficult to maintain the appropriate BG values recommended for inpatients between 3.9-10 mmol/l. Both personas experienced adverse events during their admission, which is not uncommon for inpatients. This shows the potential for a decision-support system to aid BG regulation. The personas also introduce new features for the decision-support system and show how the system can aid the current workflow for internists. A key strength of this study is that the personas have been created based on real patient stories. A key limitation is that the personas were created by one single researcher instead of a group of researchers. A recommendation is to have a validation meeting where multiple internists can help validate the personas to make them more robust. It is recommended that the final features of the decision-support system are mirrored against these personas to determine if and where the tool can have a positive impact.

4.2 Introduction

In previous chapters, it was made clear that there is a challenge in BG regulation in hospitals on a population level. Besides this, numerous obstacles that hinder doctors from effectively addressing these challenges today were introduced. Another important stakeholder in this process is the patient. Although the tool is primarily designed for internists, it is also important to include the patient in the requirement analysis.

Including patient stories and perspectives can help to understand stakeholders' characteristics better, accommodate the system to different patient needs, increase empathy among users, and make the results of this report reality-centered and tangible (Karolita et al., 2023). Future challenges on how decision-support will relate to a patient's care need can be tackled in this way.

Personas are a tool that can be utilized to obtain the patient's perspective. Personas serve as fictional representations of patient groups that can serve as a primary source of information when developing the decision-support system (Karolita et al., 2023). In the context of this study, personas are modeled by their characteristics, behavior, and ways of thinking. If a patient group has similar characteristics, behavior, and ways of thinking, they can be modeled into a single persona. This leads to the following objective for this chapter:

Provide a reality-centered exploration of BG regulation in-hospital from a patient perspective using patient personas

These personas will be developed through patient case studies. Case studies provide an in-depth exploration of an issue in a real-life setting (Crowe et al., 2011). The case studies will be formed into fictional patient personas, which are thus not traceable to the actual patients on which the personas are based.

To make the connection with a potential decision support system, the possible implementation of this system in the current workflow is highlighted in the persona stories. In this report, no extensive design is made for this tool. For now, the decision support system is described as a dashboard in which consulting internists can see the data needed for policy making, as described in Chapter 2. Besides this, the decision support system contains CGM data with predictions, as introduced in Chapter 1.

4.3 Methods

To gain a clearer understanding of the context and processes from a new perspective compared to earlier sections in this report, two personas are modeled. The number of personas two, was chosen in order to model two important groups of patients for the hospital context. These two important groups were patients who could self-regulate and require little to no help in BG regulation and patients who could not self-regulate and therefore did need a lot of help with BG regulation. No other important patient group was identified for this research. Data was collected during clinical rounds for patients who needed extra consultation regarding their BG regulation during admission. Patients from the ZGT hospital in Almelo were included in the personas. Furthermore, for inclusion, their admission had to last more than 2 days, and the case had to be identified as interesting by the researcher to offer deeper insights into the context of the decision-support tool. Data was collected between the 8th of January 2024 and the 9th of December 2024. The researcher had conversations with the patients to gather data on

persona characteristics, behavior, and ways of thinking. Besides this, the patient's EHR was checked for data reported by other medical staff. Access to a patient's EHR was granted because the researcher was also involved in their BG regulation in the hospital. The data collected from each patient is presented in Table 22.

Table 22 Collected data and definition during admission case studies.

Data collected	Discription
Reason of admission	-
Admission department	-
Relevant medical past history	Relevant to diabetes. This includes type of diabetes, micro- and macrovascular complications of diabetes, and cardiovascular disease
Medication for diabetes	-
Corticosteroid medication	-
General current history information provided by their primary doctor	Physical and emotional state of the patient, nutritional status, relevant conversation. Access to this data relied on the availability within the EHR.
General policy information provided by their primary doctor	Treatment information that is relevant to the case study. This means for example changes in relevant medication or nutritional status. Access to this data relied on the availability within the EHR.
BG regulation current history information	Status of the patient regarding BG regulation. This information was collected by the consulting doctors.
BG regulation policy information	Changes to diabetes medication made by the consulting doctors in order to regulate BG.
Re-admission instances	Including reasons of re-admission

From this data, personas were developed by a single researcher. To make the persona tangible, the developed personas include a short description of their characteristics, behavior, and admission situation. All patient information is anonymized, and traceable patient characteristics are changed. As a result of the anonymization process, the patients were not asked for informed consent to participate in this study.

4.4 Results

4.4.1 Persona 1: Scannie Sweet

Scannie is a 45-year-old female with long-existing T2D with microvascular complications. For her T2D, she used a basal-bolus regimen along with one oral medication to better control her BG levels. She is very able to check and control her BG

with her CGM sensor. She is aware of moments when she needs to prevent hypoglycemia. For safety, she always keeps her own BG values in a range between 6 and 15 at home.

Scannie was admitted to the surgery department. She was admitted for deterioration of ulcers on the right foot. She was in great pain due to this deterioration, for which she received heavy pain-relieving medication throughout her admission. The surgery department issued a visit by the internist, which is standard procedure when a person with diabetes is admitted to the hospital. The internist made a visit to talk with Scannie about her BG regulation during admission.

Even though her ulcers, the hospital context, and her pain made BG regulation more complicated, Scannie indicated that she would like to continue regulating her own BG. She was deemed competent by the internist because she was able to indicate her BG targets, between 6 and 15, like at home, she is aware when a hypoglycemia episode is coming, and she was able to explain how to change her insulin regimen according to higher or lower values. She could maintain her usual meal pattern for nutrition, simplifying her regulation process. Besides this, she had a CGM sensor with which she could get insight into her own BG data whenever she desired.

In the current workflow, it was indicated to Scannie that it was fine for her to self-regulate by the consulting internist; it remains her choice. However, it was discussed that BG targets during admission should be between 3.9-10.0 mmol/l and what the impact of higher and lower BG values can be. To make sure BG values did not rise above her target, the internist decided to check her reported BG values daily. The internist also offered help in a scenario in which Scannie did not know how to get to those target values. She could indicate to the nursing staff or her chief practitioner that she wants help.

In a workflow in which the decision-support system can be integrated, the tool with predictive features could be implemented for this patient as she has CGM data available. Personalized supervision would be issued for Scannie as the consulting internist can be alerted by the decision-support system if she would receive different nutrition and corticosteroid medication. This would mean less need for daily supervision. In addition, if Scannie decides to ask for help, the tool will aid with swift advice from the internist because accessing the right data and predictions would be easier.

The admission period was extensive. The pain and wound in the foot were reduced slowly. The discomfort experienced by Scannie made her nutritional intake level slightly lower than normal. This meant that her BG values were between 8-14 mmol/l during the rest of her admission. After 8 days, she was discharged from the hospital and could continue the last bit of wound care at home.

This persona illustrates how tolerated hyperglycemia, although it was self-regulated, can lead to slower wound healing and, therefore, extended length of stay. The persona also shows that even though a person can self-regulate, it does not mean that no supervision is needed. It still requires time and effort from internists. The decision-support system could reduce this time and effort by, for example, setting the right alarms for when high-risk situations occur or even for when certain targets are exceeded for a longer amount of time.

4.4.2 Persona 2: Robin Suline

Robin is a 71-year-old male with T2D and cardiovascular disease. In daily life, Robin treated his diabetes as a side issue, which led to overall high blood sugar levels. To counteract this, he has multiple BG-lowering oral medications combined with a basal-bolus regime at home. As a result of his attitude towards T2D, Robin did not measure his BG often at home, even though he had a CGM sensor. When he measured, he saw very high BG values, far exceeding the threshold of hyperglycemia. This had a demotivational effect on him.

One day, Robin was admitted for acute abdominal pain. A strand ileus was diagnosed. To remove this, a conventional adhesiolysis procedure was carried out successfully on the first day of his admission. Robin's BG-lowering medication was stopped for the procedure as he needed to be sober. After the procedure, Robin's state was variable, with good moments and moments of great pain. His BG-lowering medication was slowly re-introduced based on his nutritional status.

Nutrition was complicated for Robin, as his oral intake was significantly reduced in the days after the procedure due to nausea and vomiting. He received a nasogastric tube. Some intake was transferred through the nasogastric tube, although frequent vomiting prevented most of it. This caused BG values to stay between 5 and 15. Robin was feeling very sick and was unable to self-regulate. BG regulation was taken over by the internists because of this. Because of the vomiting and low intake via the nasogastric tube, BG levels were between 5 and 11. In the current workflow, the internist will from now on check the patient's status and make BG regulation policy once a day.

The next day, he also got severe diarrhea. Nutrition was changed from nasogastric tube to total parenteral nutrition with a central venous catheter. This prevented the BG levels from falling below 18 mmol/l. Unfortunately, due to staff shortage, no internist was available to help Robin with his BG regulation. The BG values decreased only after the diarrhea lessened after two days and oral nutrition began to improve. When total parenteral nutrition ceased, blood glucose levels sharply decreased to between 3.7 and 10. mmol/l. After a few days with BG values like this, Robin started to feel better, and his abdominal situation improved. This meant he could be discharged.

This persona illustrates how the hospital environment can affect persons with diabetes who cannot self-regulate. Varying situations in sickness, nutrition, and procedures result in difficult and high-risk situations. A decision-support system could aid in tracking these different situations and indicate when a situation causes a higher risk for either hypo- or hyperglycemia. In case of a procedure, the patient would have to be enrolled in the system by the chief practitioner before the procedure to be able to give an alarm to stop BG-lowering medication. Moreover, the shortage of staff during a critical period of potential deregulation, particularly with intravenous nutrition, highlighted the need for a decision-support system to assist the consulting internists in making swift decisions or alerting them of high-risk situations like Robin's. In that case, when staff shortage is a problem, quick decisions can still be made that would have benefitted Robin.

4.5 Discussion

To gain deeper insights into a key stakeholder, the patient, personas were created during the development of a decision-support system. The study shows the potential of creating personas to give context to the developers and care staff to develop and implement the decision-support system. The insight into the patient's situation can create empathy when clinicians interact with the patient and help identify potential user-system interactions in the current workflow.

The personas illustrate that it is difficult to maintain the appropriate BG values between 3.9-10 mmol/l recommended for inpatients (Pérez et al., 2020b). They also illustrate how hard it can sometimes be for clinicians to assist with BG regulation due to staff shortages. Both personas experienced adverse events during their admission, which is not uncommon for inpatients with diabetes (Moghissi et al., 2009b; Pérez et al., 2020b). This shows the potential for a decision-support system to aid BG regulation.

The two personas show the difference between a person with diabetes who can self-regulate and a person who, due to his circumstances, is not able to self-regulate. What stands out is that although a person can self-regulate, internists often still must spend time supervising the patient. This is often active supervision where, every day, a person's situation needs to be checked. In this case, a decision-support system could reduce time spent and personalize care, aiding in more passive supervision because alarms can be set to alert the internist when a situation becomes risky. Another key insight is that self-regulation is not necessarily better regulation. That depends on the patient's own BG targets. A person with diabetes can tolerate hyperglycemia themselves.

Something to highlight is the implementation of the decision-support system in the current workflow. These personas and their story include a representation of the current workflow of BG regulation by internists in the hospital. This representation shows that the decision-support system can make the workflow less time-consuming, easier, and more personalized. Due to the role of the system in identifying high-risk situations, the

system makes BG regulation care more personalized as the patients with high needs can be helped first, and patients who do not need much help can be given more freedom. The overview of data and the predictive feature that the system gives can save time and make policy-making simpler, which also aids in situations of staff shortages. These features can be implemented in the workflow at the same time that policy-making happens in the current workflow.

A key strength of this study is that the personas have been created based on real patient stories. Although these personal insights can be valuable, there are some limitations to consider. The personas were created by one single researcher instead of a group of researchers, which is more common in this type of research (Karolina et al., 2023). A recommendation is to have a validation meeting in which multiple internists and patients can help validate the personas to make them more robust. If this validation study indicates that the two personas presented do not adequately represent all of this stakeholder's types, it's advisable to conduct a workshop to generate additional personas. It is recommended that the final features of the decision-support tool are mirrored to these personas to determine if and where the tool can create a positive impact.

4.6 Conclusion

The personas from this study help understand patient characteristics better, accommodate the system to different patient needs, increase user empathy, and make the report's results reality-centered and tangible. The personas show that it is difficult to maintain appropriate BG values during admission, which can lead to adverse events. Besides this, the added insights and alarms can help internists prioritize patient needs in staff shortage situations, making the current workflow more effective. To be able to use these personas optimally, they should be validated by multiple researchers. If the personas accurately reflect different patient types, they can have a positive impact on the further development and implementation of the decision-support system.

5. Final report conclusions & a proposal for a simple decision-support system

5.1 Conclusions

The report as a whole aims to provide a comprehensive analysis of the challenges related to BG regulation in hospitalized persons with diabetes and to develop a decision-support system to improve clinical management and optimize patient outcomes. The three chapters leading up to this conclusion treated different aspects of obtaining this objective. The first chapter explored patient-specific ARIMA(X) models in an adaptive identification framework to predict BG values with short-term PHs in admitted persons with diabetes. The second chapter explored key obstacles faced by doctors when trying to achieve tight glucose regulation in admitted patients using qualitative data. Besides this, the doctor's methods for achieving this tight regulation were analyzed. The third chapter provided a reality-centered exploration of BG regulation for admitted patients with diabetes through patient personas. In this final conclusion, the aspects that can aid the development of a decision-support system right now are listed and combined into a simple decision-support system proposal.

The most relevant insight for current use from chapter one is that the models developed can predict BG values with a PH of 30 minutes in a clinically relevant way based on only CGM data during admission. Preferably, data pre-admission is also available so that the model can be trained on that data and can be used immediately. Although PH=30 for the intra- and cross-BG data does not satisfy the CEGA threshold of 100% in this study, it can still be seen as clinically relevant as the ISO standard is 99% in A+B zones. Both input data combinations do satisfy that threshold. This possible predictive feature enables medical staff to take proactive action based on a 30-minute-ahead warning that a hypo- or hyperglycemic event is coming up. Both BG-lowering and rising actions, like fast-acting insulin or carbohydrates, can avert adverse events in such a scenario (Prendin et al., 2021)(Spanakis et al., 2018).

This proactive action can counteract one of the challenges perceived by doctors in chapter two. Doctors perceive add-on insulin schemes as causing delays in BG regulation as they must act on BG values from the past. The decision-support system with 30-minute-ahead warnings could replace add-on insulin schemes in an ideal scenario. Chapter two reveals another relevant challenge that should be addressed by the decision-support system. When trying to get a complete view of an admitted patient, doctors perceive a challenge because they do not have all the relevant data available in a simple yet comprehensive way. This is especially the case for nutritional status and home diabetes medication. The integration of this data with other relevant data for policy determination was proposed as a solution by the doctors interviewed.

In case of staff shortages, this data should especially be available for patients with high risk, for example, patients with enteral or parenteral nutrition and/or corticosteroid medication as identified by chapter 2, 3 and Moghissi et al. (Moghissi et al., 2009b). The extra insight into high-risk situations, as well as into upcoming hypo- or hyperglycemic events, could then, in turn, help doctors give patients extra insights into their regulations.

To summarize, with the results from this report, a decision-support system that could currently be implemented for testing should contain an overview of data perceived as essential, which is sometimes hard to get by. It should contain CGM data with a 30-minute-ahead warning for hypo- or hyperglycemia as a predictive feature. It should also contain a way to identify high-risk patients based on nutritional and/or medication data.

5.2 Proposal for a simple decision-support system

This report aimed to develop a decision-support system to improve clinical management of BG regulation and clinical outcomes. It is important to realize that the outcomes of this report, combined with the literature, can only lead to a simple proposal, which needs to be tested in a hospital setting and undergo a lot of development. However, development can be started soon for a simple but scientifically based decision-support system for the ZGT hospital in Almelo.

The simple overview of essential data is the first step in this system. This could be implemented in the form of a dashboard. Essential data contained in this dashboard should be current and past BG values, home & admission diabetes medication, corticosteroid medication, nutritional status, and admission reason. Additional information could be a recent HbA1c lab value, self-regulation ability, medical past history, and insulin-carbohydrate ratio. All this information, except for self-regulation and insulin-carbohydrate ratio, is contained in the EHR of the ZGT. To make sure this information is retrieved from EHR and put into a comprehensive overview, the ZGT database can be used. The ZGT data department makes emergency copies from the EHR every ten minutes. An emergency copy contains all patient information needed in a 10-minute delayed version. From this copy, the essential information can be copied into a dashboard that is made using the R shiny package in R. This dashboard could be used internally in the ZGT to supply the internists with a comprehensive overview of people with diabetes. A notification can be integrated into this dashboard if enteral or parenteral nutrition or corticosteroid medication is detected to notify doctors of high-risk scenarios. A limitation of this system is that it would not contain information on self-regulation or insulin-carbohydrate ratio, as that information is contained in either conversation or insulin pumps that cannot be connected as easily to this dashboard.

A slightly more complicated step is to integrate real-time CGM with a 30-minute-ahead warning into the dashboard and system. The BG values described in the previous

paragraph are measured by finger stick or CGM measurement and put into the EHR approximately four times per day. Real-time CGM values from sensors are, unfortunately, not connected to the EHR. Retrieving real-time CGM data is not an easy task. For now, the most viable option is to use the SATO BANI, a device developed by researchers in the ZGT to transfer real-time CGM values to the ZGT database (Reference not available yet). A large limitation is that this device is not yet ready for implementation. A new researcher with the right expertise needs to be found to finish the device.

Provided that real-time CGM data can be retrieved from these sensors, the proposed ARIMA(X) with adaptive identification framework can be applied to this data and can be run on the ZGT research servers in real-time. This model can be used to set alarms when the model's prediction indicates that BG values exceed either hypo- or hyperglycemia thresholds. The upper limit for BG values is defined as 10 mmol/l and the lower limit as 5.6 mmol/l to minimize the risk of hypoglycemia (Moghissi et al., 2009b; Pérez et al., 2020b). If upcoming hyperglycemia is detected, proactive action is initiated, according to the study by Perez et al. (Pérez et al., 2020b). Perez et al. developed an algorithm for the initiation of insulin in noncritically ill hyperglycemic patients in a hospital. The patients that can be treated by this algorithm are both T1D and T2D with a basal-bolus regimen. The algorithm can recommend a correctional insulin bolus based on the total daily dose (TDD) of insulin or weight. In the study by Perez et al., bolus doses are recommended based on premeal glucose. The decision support system should use these doses for predicted BG values instead of premeal BG. All recommended insulin bolus amounts, as reported by Perez et al. for different premeal BG values, are reported in Table 23. In this table, however, premeal BG is swapped for predicted BG. If a hypoglycemic alarm is raised, Urbanova et al. recommend that a 15-gram dose of oral glucose (in the desired form) be given (Urbanová et al., 2022). Fifteen minutes after this dose, CGM values should be checked again. If BG values are still insufficient, 15 grams of oral glucose should be repeated. It should be noted that this glucose amount is from a review on the effectiveness of carbohydrate treatment in nonsevere hypoglycemia in adults with insulin-treated diabetes, so this was not tested in a hospital environment. The advice for these different doses based on predicted BG values will be incorporated inside the dashboard as a separate section to keep a comprehensive overview.

*Table 23 correctional insulin boluses for predicted BG values. These values are predicted by the ARIMA(X) with an adaptive identification algorithm. The doses are based on TDD or weight. *If scheduled premeal bolus (Pérez et al., 2020b)*

Predicted BG values (mmol/l)	**<60 units of insulin/day or <60 kg**	**40-80 units of insulin/day or 60-90 kg**	**>80 units of insulin or >90 kg**
< 4.4	-1*	-1*	-2*
4.5-7.8	0	0	0
7.9-11.1	+1	+1	+2

11.2-13.9	+2	+3	+4
14.0-16.7	+3	+5	+7
>16.8	+4	+7	+10

Future recommendations for the decision-support system can be found in the different chapters, and more research is needed before implementation.

Bibliography

Boiroux, D., Aradóttir, T. B., Nørgaard, K., Poulsen, N. K., Madsen, H., & Jørgensen, J. B. (2017). An Adaptive Nonlinear Basal-Bolus Calculator for Patients with Type 1 Diabetes. *Journal of Diabetes Science and Technology*, *11*(1), 29–36. https://doi.org/10.1177/1932296816666295

Boulanger, J.-L. (2017). *Certifiable Software Applications 3. Chapter 2 Requirements Management*.

Clarke, W. L., Cox, D., Gonder-Frederick, L. A., Carter, W., & Pohl, S. L. (n.d.). Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. In *DIABETES CARE* (Vol. 10, Issue 5). http://diabetesjournals.org/care/article-pdf/10/5/622/515265/10-5-622.pdf

Crecil Dias, C., Kamath, S., & Vidyasagar, S. (2020). Blood glucose regulation and control of insulin and glucagon infusion using single model predictive control for type 1 diabetes mellitus. *IET Systems Biology*, *14*(3), 133–146. https://doi.org/10.1049/iet-syb.2019.0101

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, *11*. https://doi.org/10.1186/1471-2288-11-100

Dong, Y., & Peng, C.-Y. J. (2013). *Principled missing data methods for researchers*. http://www.springerplus.com/content/2/1/222

Fortin, A., Rabasa-Lhoret, R., Roy-Fleming, A., Desjardins, K., Brazeau, A. S., Ladouceur, M., & Gingras, V. (2017). Practices, perceptions and expectations for carbohydrate counting in patients with type 1 diabetes – Results from an online survey. *Diabetes Research and Clinical Practice*, *126*, 214–221. https://doi.org/10.1016/j.diabres.2017.02.022

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health*, *25*(10), 1229–1245. https://doi.org/10.1080/08870440903194015

Freckmann, G., Pleus, S., Grady, M., Setford, S., & Levy, B. (2019). Measures of Accuracy for Continuous Glucose Monitoring and Blood Glucose Monitoring Devices. In *Journal of Diabetes Science and Technology* (Vol. 13, Issue 3, pp. 575–583). SAGE Publications Inc. https://doi.org/10.1177/1932296818812062

Hyndman, R. J. (2007). *Minimum sample size requirements for seasonal forecasting models*.

Karolita, D., McIntosh, J., Kanij, T., Grundy, J., & Obie, H. O. (2023). Use of personas in Requirements Engineering: A systematic mapping study. In *Information and Software Technology* (Vol. 162). Elsevier B.V. https://doi.org/10.1016/j.infsof.2023.107264

Kovatchev, B. P., Patek, S. D., Ortiz, E. A., & Breton, M. D. (2015). Assessing sensor accuracy for non-adjunct use of continuous glucose monitoring. *Diabetes Technology and Therapeutics*, *17*(3), 177–186. https://doi.org/10.1089/dia.2014.0272

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). How sure are we that economic time series have a unit root?*. In *Journal of Econometrics* (Vol. 54).

Laguna, A. J., Rossetti, P., Ampudia-Blasco, F. J., Vehí, J., & Bondia, J. (2014). Identification of intra-patient variability in the postprandial response of patients with type 1 diabetes. *Biomedical Signal Processing and Control*, *12*(1), 39–46. https://doi.org/10.1016/j.bspc.2013.07.003

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Moghissi, E. S., Korytkowski, M. T., DiNardo, M., Einhorn, D., Hellman, R., Hirsch, I. B., Inzucchi, S. E., Ismail-Beigi, F., Kirkman, M. S., & Umpierrez, G. E. (2009a). American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control. *Diabetes Care*, *32*(6), 1119–1131. https://doi.org/10.2337/dc09-9029

Moghissi, E. S., Korytkowski, M. T., DiNardo, M., Einhorn, D., Hellman, R., Hirsch, I. B., Inzucchi, S. E., Ismail-Beigi, F., Kirkman, M. S., & Umpierrez, G. E. (2009b). American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control. *Diabetes Care*, *32*(6), 1119–1131. https://doi.org/10.2337/dc09-9029

Mohebbi, A., Johansen, A. R., Hansen, N., Christensen, P. E., Tarp, J. M., Jensen, M. L., Bengtsson, H., & Mørup, M. (2020). *Short Term Blood Glucose Prediction based on Continuous Glucose Monitoring Data*. http://arxiv.org/abs/2002.02805

Oluwaseye Joel, L., Doorsamy, W., & Sena Paul, B. (2022). A Review of Missing Data Handling Techniques for Machine Learning. *International Journal of Innovative Technology and Interdisciplinary Sciences Www.IJITIS.Org*, *5*(3), 971–1005. https://doi.org/10.15157/IJITIS.2022.5.3.971-1005

Oviedo, S., Vehí, J., Calm, R., & Armengol, J. (2017). A review of personalized blood glucose prediction strategies for T1DM patients. *International Journal for Numerical Methods in Biomedical Engineering*, *33*(6), 1–21. https://doi.org/10.1002/cnm.2833

Pérez, A., Ramos, A., & Carreras, G. (2020a). Insulin Therapy in Hospitalized Patients. *American Journal of Therapeutics*, *27*(1), E71–E78. https://doi.org/10.1097/MJT.0000000000001078

Pérez, A., Ramos, A., & Carreras, G. (2020b). Insulin Therapy in Hospitalized Patients. *American Journal of Therapeutics*, *27*(1), E71–E78. https://doi.org/10.1097/MJT.0000000000001078

Phadke, R., Prasad, V., Nagaraj, H. C., & Bhograj, A. (2020). Univariate data-driven models for glucose level prediction of CGM sensor dataset for T1DM management. *Sādhanā*, *45*. https://doi.org/10.1007/s12046-020-1277-8S

Prendin, F., Del Favero, S., Vettoretti, M., Sparacino, G., & Facchinetti, A. (2021). Forecasting of glucose levels and hypoglycemic events: Head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only. *Sensors*, *21*(5), 1–21. https://doi.org/10.3390/s21051647

Prendin, F., Díez, J. L., Del Favero, S., Sparacino, G., Facchinetti, A., & Bondia, J. (2022). Assessment of Seasonal Stochastic Local Models for Glucose Prediction without Meal Size Information under Free-Living Conditions. *Sensors*, *22*(22). https://doi.org/10.3390/s22228682

Rousseau, M. P., Beauchesne, M. F., Naud, A. S., Leblond, J., Cossette, B., Lanthier, L., Grondin, F., & Bernier, F. (2014). An interprofessional qualitative study of barriers and potential solutions for the safe use of insulin in the hospital setting. *Canadian Journal of Diabetes*, *38*(2), 85–89. https://doi.org/10.1016/j.jcjd.2014.01.013

Sawaryn, B. (2020). *MODELLING FLUCTUATIONS OF BLOOD GLUCOSE LEVELS BASED ON FOOD INTAKE AND PHYSICAL ACTIVITY IN PATIENTS WITH DIABETES MELLITUS TYPE II Development of a controlled protocol which administers perturbations in daily life situations*.

Selim, M., Zhou, R., Feng, W., & Alam, O. (2020). Reducing error propagation for long term energy forecasting using multivariate prediction. *EPiC Series in Computing*, *69*, 161–169. https://doi.org/10.29007/mbb7

Smith, G. J., Abraham, M. B., De Bock, M., Fairchild, J., King, B., Ambler, G. R., Cameron, F., Mcauley, S. A., Keech, A. C., Jenkins, A., Davis, E. A., O'neal, D. N., & Jones, T. W. (2023). Impact of Missing Data on the Accuracy of Glucose Metrics from Continuous Glucose Monitoring Assessed Over a 2-Week Period. *Diabetes Technology and Therapeutics*, *25*(5), 356–362. https://doi.org/10.1089/dia.2022.0101

Spanakis, E. K., Levitt, D. L., Siddiqui, T., Singh, L. G., Pinault, L., Sorkin, J., Umpierrez, G. E., & Fink, J. C. (2018). The Effect of Continuous Glucose Monitoring in Preventing Inpatient Hypoglycemia in General Wards: The Glucose Telemetry System. *Journal*

of Diabetes Science and Technology, *12*(1), 20–25. https://doi.org/10.1177/1932296817748964

Spanakis, E. K., Urrutia, A., Galindo, R. J., Vellanki, P., Migdal, A. L., Davis, G., Fayfman, M., Idrees, T., Pasquel, F. J., Coronado, W. Z., Albury, B., Moreno, E., Singh, L. G., Marcano, I., Lizama, S., Gothong, C., Munir, K., Chesney, C., Maguire, R., … Umpierrez, G. E. (2022). Continuous Glucose Monitoring–Guided Insulin Administration in Hospitalized Patients With Diabetes: A Randomized Clinical Trial. *Diabetes Care*, *45*(10), 2369–2375. https://doi.org/10.2337/dc22-0716

Summary of product characteristics: NovoRapid. (n.d.).

Urbanová, J., Frier, B. M., Taniwall, A., Brožová, K., Malinovská, J., Chandel, A., & Brož, J. (2022). Optimal Carbohydrate Dose for Treatment of Nonsevere Hypoglycemia in Insulin-Treated Patients With Diabetes: A Narrative Review. In *Canadian Journal of Diabetes* (Vol. 46, Issue 7, pp. 743-749.e4). Elsevier B.V. https://doi.org/10.1016/j.jcjd.2022.03.011

Yang, J. (2019). *An_ARIMA_Model_With_Adaptive_Orders_for_Predicting_Blood_Glucose_Concentrations_and_Hypoglycemia*.

Zelada, H., Perez-Guzman, M. C., Chernavvsky, D. R., & Galindo, R. J. (2023). Continuous glucose monitoring for inpatient diabetes management: an update on current evidence and practice. In *Endocrine Connections* (Vol. 12, Issue 10). BioScientifica Ltd. https://doi.org/10.1530/EC-23-0180

Appendix:

Literature ARIMA(X) modelling for BG prediction

Phadke et al. evaluated the performance of an ARIMA model on two datasets, the Libre Pro CGM sensor dataset and the Ohio T1DM dataset(Phadke et al., 2020). The Libre Pro CGM dataset contained data from 10 type 1 diabetes patients measured with Abbott's Flash Glucose Monitoring system, which measures glucose every 15 minutes for 14 days. The Ohio T1DM dataset contained glucose data of 6 patients with type 1 diabetes measured with a Minimed Enlite glucose sensor, which measures every 5 minutes. Measurements were collected over an eight-week period. As pre-processing, missing data was interpolated. An overlapping forward window method is used for training and testing with a window of 12 hours. An ARIMA(2,1,0) was applied without specification on the rationale behind this choice of model. The performance metrics for this study can be seen in Table 1.

Prendin et al. did two studies(Prendin et al., 2021, 2022). The first study compares linear and non-linear data-driven algorithms based on only CGM data in type 1 diabetes patients (Prendin et al., 2021). All algorithms were tested on the same dataset of 124 CGM traces selected from 141 patients. The traces contained data from patients who wore the Dexcom G6 sensor for 10 consecutive days. The sensor had a sampling time of 5 minutes. 5.4% of the total samples were below 3.9 mmol/l. For the purpose of this report, this section will focus on the testing of the individualized linear black-box models, which were AR, ARMA, and ARIMA models. It is worth noting that according to this study, non-linear models do not significantly outperform linear models and that individualized models outperform population ones. Model parameters for the linear models were based on BIC criteria. The identification happens among AR values = {1,...,30}, MA = {1,...,15}. Parameter estimation was done by either the prediction error method or recursive least-squares. Recursive linear least squares fall into a fixed structure but time-varying parameters approach. Only stable models were identified. Predictions with a 30-minute PH using a Kalman filter framework. Root mean square error (RMSE) was utilized as a measure for accuracy.

The second study by Prendin et al. assessed a newly proposed method of seasonal stochastic local models for glucose prediction against other linear and nonlinear models under free-living conditions (Prendin et al., 2022). All the models from this study were trained and tested based on the Ohio T1DM dataset and the CTR3 dataset. The Ohio T1D dataset is previously described. The CTR3 dataset contains data from 14 individuals with type 1 diabetes with hybrid closed-loop insulin systems. Glucose was monitored with a Dexcom G4 sensor, and meals and insulin doses were recorded. 6-to 7 weeks of data were used for training, and 10 days were reserved for testing. The models were all trained on individual patient data, and parameters were identified using the

Bayesian information criterion (BIC). The identification happens among AR values = {1,...,20}, MA = {1,...,20}, I = {0,1} and X = {1,...20}. These models were evaluated on accuracy using RMSE for PH of 30, 45, 60, and 75 minutes. As before, this section will focus on ARIMA and ARIMAX models used as a comparison in this study. The newly proposed method clusters PP periods and applies a seasonal form of ARIMA (SARIMA). Notably, this method only outperformed ARIMA and ARIMAX for post-prandial periods and higher PH of 60 and 75 minutes.

A study by Mohebbi et al. studies recurrent neural networks based on population and individual patient data (Mohebbi et al., 2020). It uses ARIMA models on the same data as a comparison. CGM data was acquired from 50 patients for 14 consecutive days with a monitoring frequency of 5 minutes. Only CGM days and periods with more than 70% available readings are included. Missing data is further interpolated linearly. No information exists about the type of diabetes or treatment regimens. Fifteen patients are used for the development of the patient-specific models. The data of these patients is divided into seven days of training data and seven days of testing data. The training data is further partitioned into 1, 3, and 7 days to examine whether increasingly added CGM data will have an impact on model performance. As preprocessing steps, data has been normalized to have zero mean and a variance of one. The identification of parameters happens among AR values = {1,...,24}, MA = {1,...,24}, I = {0,5} based on the AIC criterium. PH of 15 30, 45, 60, and 90 minutes are tested. Performance is tested with MAE and RMSE. The study concludes that LTSM is the best-performing model with seven days of training data, however, with little training data, the LTSM does not perform well, while the performance of the ARIMA model decreases less.

Ben Saweryn wrote a master's thesis on modeling fluctuations of blood glucose levels based on food intake and physical activity in patients with diabetes mellitus type 2 (Sawaryn, 2020). Patients-specific AR and ARMAX models to analyze the effect of lifestyle interventions. The study included five patients in semi free-living conditions. A food protocol managed their intake for the duration of the study. BG values were measured with a Freestyle Libre glucose sensor that takes a sample every 15 minutes. Besides this, food intake in carbohydrates and fats (g) and steps per minute were measured by other means. Missing data was linearly interpolated, and the full data array was resampled to 15 minutes for uniformity. The identification of parameters happens among AR values = {1,...,10}, and for the ARMAX model MA = {1,...,10} and X = {1,...,10} based on the AIC criterium. For simplicity, all parameters were set to the same amount. Moreover, the training set was divided into three parts; models with multiple parameters were developed for each part, and the AIC values were averaged between them. The models that showed the minimum mean AIC were considered to have the optimal combination of parameters. Furthermore, for X, a delay term was included, Nk = {0,...,2}, to model the delay of the effects for each exogenous input. The final model was trained on the complete training set. The training set varied between 4 to 12 days in increments

of one day, with a test set also varying in length with the variation of training data. Performance of the model was measured using RMSE for PH 15, 20, 60, 120, and 180 minutes ahead. Due to the objective of this study, the models with the lowest RMSE at a PH of 120 minutes with a set training size were analyzed further with Clarke's error grid analysis (CEGA). For each subject, the RMSE with optimal training set sizes and optimal model parameters are displayed in this study. For the purpose of comparison, all RMSEs were averaged in Table 1. The mean CEGA values for all patients with optimal training set size are displayed in Table 1. This research shows the need for patient-specific models as the number of parameters and optimal training set size varies among patients.

Yang et al. conducted research with the objective of showing the added value of an ARIMA model with adaptive orders(Yang, 2019). CGM is measured by a Glutalor CGM DS-02 sensor, which measures a BG value every 3 minutes. Training is done using a database of 5 T1D and 5 T2D patients in free living conditions. Before applying the model, the supply evidence that stationarity of CGM data changes over time with augmented Dickey-Fuller tests (ADF) and analysis of variance (ANOVA). CGM data gets stationary over time. To make the model, a window of length Nw of 400 samples is chosen and determined by ADF tests. The identification of parameters happens among AR values = {1,...,7}, MA = {1,...,3}, based on the AIC criterium. Parameter values are estimated using least squares estimation. Predictions are made with PH = 30. After a prediction step, the process is started again from the AIC criteria to determine new parameters. Model performance is assessed using relative absolute deviation, the sum of squares of prediction error, and the average normalized temporal gain. They show that this model outperforms an ARIMA model with comparable parameters but without the dynamic element.

Interview questions

Introduction:

- Context: product development program in Porto
- Context: research on challenges in BG regulation in the hospital for admitted people with diabetes.

Questions:

1. What are your biggest challenges in regulating BG values for admitted people with diabetes?

2. How are decisions about BG regulation made at the moment, and which tools or systems do you use for that?
3. Do you currently use continuous glucose monitors or other technologies to monitor BG values? How do they integrate with the rest of your hospitals systems?
4. What are long term goals for your hospital concerning BG management and/or the application of technology in this field?

5. Did I miss something crucial or is there something you want to add concerning this topic?