Stochastic Model Predictive (STOMP) Glycaemic Control for the Intensive Care Unit: Development and virtual trial validation

Kent W. Stewart¹, Christopher G. Pretty¹, Hamish Tomlinson¹, Liam Fisk¹, Geoffrey M. Shaw², J. Geoffrey Chase¹.

¹Department of Mechanical Engineering, Centre for Bio-Engineering, University of Canterbury, Christchurch, New Zealand.

²Department of Intensive Care, Christchurch Hospital, Christchurch, New Zealand.

ABSTRACT

Critically ill patients often experience stress-induced hyperglycaemia, which results in increased morbidity and mortality. Glycaemic control (GC) can be implemented in the intensive care unit (ICU) to safely manage hyperglycaemia. Two protocols SPRINT and STAR, have been implemented in the Christchurch ICU, and have been successful in treating hyperglycaemia while decreasing the risk of hypoglycaemia. This paper presents a new GC protocol that implements the probabilistic, stochastic forecasting methods of STAR, while formalizing the control methodology using model predictive control (MPC) theory to improve the ability to tune the dynamic response of the controller. This Stochastic Model Predictive (STOMP) controller predicts the response to a given insulin/nutrition intervention, and attributes weighted penalty values to several key performance metrics. The controller thus chooses an intervention at each hour that minimizes the sum of these penalties over a prediction window of 6 hours, which is twice as long as the 3-hour window used in STAR. Clinically validated virtual trials were used to evaluate the relative performance of STOMP. Results showed STOMP was able to obtain results very similar to STAR with both protocols maintaining approximately 85% of time within 4.4-8.0 mmol/L glycaemic band, and only 4-5 patients of the 149 patient STAR cohort having blood glucose (BG) < 2.2 mmol/L. STOMP was able to attain similar results to STAR while further increasing ease of controller tuning for different clinical requirements and reducing the number of BG measurements required by 35%.
INTRODUCTION

Critically ill patients often experience stress-induced hyperglycaemia and high levels of insulin resistance [1-7]. The occurrence of hyperglycaemia, predominantly severe hyperglycaemia, is associated with an increase in morbidity and mortality in this group of patients [1, 3]. Glycaemic variability, and thus poor control, is also independently associated with an increase in mortality [8, 9].

It has been shown that effective glycaemic control (GC) can significantly reduce the number of negative outcomes associated with poor control by modulating nutrition and/or insulin administration [7, 10, 11]. Effective GC can also lead to a reduction in the rate and severity of organ failure [12] and the cost of care [13, 14]. However, consistent, safe and effective GC remains elusive with several other studies achieving negative, or inconclusive outcomes [15-20]. In addition, there is little agreement on what constitutes desirable glycaemic performance [21-23], particularly with regard to how GC affects outcome.

The model-derived SPRINT protocol has been successful at reducing organ failure and mortality [10, 12] with a patient-specific approach, providing the tightest control across all patients of several large studies [24, 25]. As a series of interactive charts, the SPRINT protocol allowed nutrition and insulin interventions to be tailored to current patient condition. However, as a paper-based protocol, SPRINT was relatively inflexible to different desired blood glucose targets and clinical uses, and required a relatively high nurse workload with 1-2 hourly blood glucose (BG) measurements.

The Stochastic TARgeted (STAR) glycaemic control protocol was thus developed to address these issues [26, 27]. STAR recommends an intervention based on a clinically specified maximum risk of mild hypoglycaemia (e.g. BG < 4.4 mmol/L), derived from stochastic model predictions of future insulin sensitivity [28, 29]. With the ability to quantify the probability of hypoglycaemia, STAR allows aggressive yet safe control of blood glucose within a target band. STAR is flexible to different blood glucose targets [30, 31] and nursing intervention frequency, and thus, addresses many of the areas for improvement with the SPRINT protocol. However, the intervention selection algorithm used by STAR is fixed and does not allow for dynamic tuning, which limits the capacity for the controller to be further optimized in real time.
Model predictive control (MPC) is an alternative control approach that allows the dynamic response of the controller to be easily tuned through a series of clinically pre-defined cost functions. MPC utilizes a mathematical model of a system to forecast the response to a given input, and control interventions are chosen to produce optimal forecasted results. Commonly, optimization will involve specifying weighted (cost) functions to key input and output performance metrics, and choosing an intervention that minimizes the sum of these values. The benefit of such a system is that the cost functions can be easily optimized to produce robust and consistent control outcomes from an intuitively easily understood clinical specification. This type of controller was chosen due to the flexibility of cost functions in allowing the dynamic response of the controller to be easily tuned. MPC has also been used for glycaemic control with a different model [32-35].

This article presents a Stochastic Model Predictive (STOMP) GC protocol that uses a low error, infrequently measured, BG signal to control the BG levels in adult ICU patients while providing greater flexibility than STAR. This research presents the protocol design and optimization for an adult ICU using clinically validated [36] virtual trials to amend safety and efficacy before clinical uptake.
METHODS

Glucose-Insulin Model

A variant of the ICING model [37] was used to describe glucose-insulin metabolic system dynamics:

\[
\begin{align*}
\dot{G}(t) &= -p_G G(t) - S_I(t)G(t) \frac{Q(t)}{1 + \alpha_G Q(t)} + \frac{P(t) + EGP - CNS}{V_G} \\
\dot{Q}(t) &= n_I(I(t) - Q(t)) - n_c \frac{Q(t)}{1 + \alpha_G Q(t)} \\
\dot{I}(t) &= -n_K I(t) - n_L \frac{I(t)}{1 + \alpha_I I(t)} - n_I(I(t) - Q(t)) + \frac{u_{ex}(t)}{V_I} + (1 - x_L) \frac{u_{en}(G)}{V_I} \\
P(t) &= \min(d_2 P_2, P_{\text{max}}) + P N(t) \\
\dot{P}_1(t) &= -d_1 P_1 + D(t) \\
\dot{P}_2(t) &= -\min(d_2 P_2, P_{\text{max}}) + d_1 P_1 \\
u_{en}(G) &= \min(\max(u_{\text{min}}, k_1 G(t) + k_2), u_{\text{max}})
\end{align*}
\]
Stochastic Model

Forward prediction of BG is enabled by an estimate of the conditional probability density function of SI based on historical data (stochastic model). The stochastic model used by STAR is generated using kernel-density methods and model-based insulin sensitivity data from a large cohort of patients (43,000 SI values from approximately 400 patients). Given a value of SI (at time \( n \)), the stochastic model can be used to estimate the probability of future SI values (at time \( n+1 \)).

![Stochastic model of insulin sensitivity.](image)

**Figure 1: Stochastic model of insulin sensitivity.**

STAR focusses on the 5\(^{th}\)- and 95\(^{th}\)-percentile values, as these values can be used to impose a 5% risk limit on hypoglycaemia for a given insulin/nutrition intervention. Figure 2 indicates the relationship between the insulin sensitivity and the associated blood glucose trajectory. The model covers a broad medical ICU cohort over all the days of stay, but can be made specific to unique cohorts [39, 40].
Using the insulin sensitivity stochastic model obtained from Figure 1, the BG stochastic model, for a specific insulin and nutrition intervention, can be obtained through solving the Glucose-Insulin model. The BG stochastic model is used to predict the trajectory of the patients BG, as seen in the right of Figure 2.

**Cost functions**

The cost functions used to evaluate the multiple different interventions can be seen in Figure 3. These cost functions were chosen iteratively using virtual trials to evaluate the likely results from an individual set of functions. They are also strongly based on clinical and physiological literature and clinical experience with prior protocols. The functions are designed to manage risk tradeoffs between BG outcomes, and the insulin and nutrition interventions.

The 5th- and 95th-percentile BG cost functions are based on BG predictions with given interventions and an associated predicted SI value from the stochastic model. These functions were designed to severely penalize both hyper- and hypo-glucemia, and thus keep the predicted 5th- and 95th-percentile BG range within a desirable band [41]. The 5th-percentile BG cost function was generated by a combination of functions, severely penalizing BG < 4 mmol/L while not forcing the controller to be too aggressive in lowering BG, ensuring hypoglycaemia was very unlikely to occur. Thus, the 95th-percentile BG cost
function is compromised to put greater emphasis on reducing the incidence of hyperglycaemia. The equations for the two cost functions can be seen in Equation (8) and (9).

\[
Cost_{BG5} = 1.875 * BG_{5th}^2 - 16.5 * BG_{5th} + 36.3 + e^{(-10*BG_{5th}+42)} \quad (8)
\]

\[
Cost_{BG95} = 0.1 * BG_{95th}^2 \quad (9)
\]

Figure 3: Cost functions used to evaluate the optimal intervention.

The insulin and nutrition cost functions were designed to maximize nutrition and minimize insulin use. This choice ensures the patient is getting as much of a desired goal nutrition rate as possible to get sufficient nutrients for recovery, while minimizing the chance of BG falling dramatically due to the large insulin doses amplifying the effects of insulin sensitivity variability. Hence, it is a balance weighted
towards nutrition intake and safety. The equations for the two cost functions can be seen in Equation (10) and (11).

\[ Cost_{INS} = 0.0011 \cdot \text{Insulin}^2 - 0.0775 \cdot \text{Insulin} + 1.255 \] \hspace{1cm} (10)

\[ Cost_{NUT} = 1.1719e4 \cdot \%GF^4 - 1.901e4 \cdot \%GF^3 + 1.1697e4 \cdot \%GF^2 - 3.2726 \cdot \%GF + 356.0714 \] \hspace{1cm} (11)

Pancreatic insulin secretion in the critically ill can be highly variable and unpredictable. Thus, maintaining a low insulin infusion rate can dominate pancreatic insulin secretion [42] and provide greater certainty around circulating plasma insulin for model-based predictions. Hence, the minimum of the insulin cost function is placed at 2 U/hr (33 mU/min) to increase this predictability.

The mathematical equations for the cost functions were derived by fitting either a 2nd or 4th order polynomial to the desired clinical shape. The cost function polynomials were designed to have a clear global minima in the range at which the functions would be applied. The use of polynomials meant the cost functions were smooth, thus meaning convergence to the global minimum was regular. An exponential term was added to the BG 5th Cost function to severely penalise hypoglycaemia.

**Cost function weightings**

Weightings were placed on each of the evaluated cost functions to increase the relative importance of certain variables and factors. The costs of the hourly BG predictions, evaluated up to the prediction horizon, were weighted so that the weighting increased as the blood glucose prediction time increased. A normalized weighting sum was maintained of this cost value, defined.

\[ W_i = \left( \frac{2}{N} - \frac{1}{\sum_{j=1}^{N} i} \right) \frac{i}{N} \] \hspace{1cm} (12)

Where: \( W_i = \) Normalized increasing weight

\( i = BG \) prediction hour (\( i = 1 \) – 6 hours)

\( N = Prediction \) window period (6 hours)
In addition to the time dependent normalized weighting of BG predictions, an additional weighting was placed on the blood glucose prediction costs. This weighting ranks the BG prediction costs higher than the insulin and nutrition costs. Iteratively, it was found that a weighting multiplier of 6 on the BG prediction costs gave the best performance, while maintaining reasonable nutrition and insulin levels with this overall model-based approach.

**Prediction horizon**

Prediction horizons of 1 to 10 hours were initially investigated. For simplicity, the nutrition and insulin interventions chosen are kept constant over the finite prediction horizon. This design choice was made to restrict the BG variability introduced by unexpected controller behaviour. Insulin and nutrition changes have differing timescales, with insulin changes having a rapidly observable effect (<10 minutes), while enteral nutrition changes act over 1-2 hours due to the slower, more complex absorption dynamics through the stomach and gut. Thus, BG outcomes over longer prediction horizons include greater contribution from changes in enteral nutrition.

Initial simulations highlighted two factors that limited the prediction horizon. Insulin sensitivity fluctuates each hour, and thus the current fitted SI value becomes more inaccurate as the prediction horizon increases. Additionally, constant insulin and nutrition over the interval means the model reaches a steady-state value over time. Typically, steady-state was reached after approximately 6 hours, thus limiting the maximum horizon. As a result, a prediction horizon of 6 hours was chosen.

To ensure all the evaluated costs of 5th- and 95th-percentile BG had equal weightings for each prediction horizon tested, relative to the other parameters, the total cost for the 5th- and 95th-percentile BG was set to the average of the costs evaluated at each hour over the prediction horizon. This summation over 6 hours effectively awards consistent BG within the band.
**Patient Cohort and virtual trials**

Clinical data from 149 patients treated with the STAR protocol (2011-2014) [26, 30], in Christchurch Hospital ICU were used to generate virtual patients. Details of these patients are shown in Table 2. The Upper South Regional Ethics Committee, New Zealand granted approval for the audit, analysis and publication of this data.

Virtual patients were created from the patient-specific time varying model-based insulin sensitivity profiles [43]. This model-based insulin sensitivity can be used as a critical marker of patient metabolic state [29, 36, 44]. These virtual patients allow robust protocols to be safely designed and rigorously tested prior to clinical implementation, improving patient safety and minimising the need for protocol alterations post-implementation [36]. Performance and safety were assessed by whole-cohort %BG in glycaemic bands, and the %BG and number of patients with hypoglycaemia.

**Table 2: STAR cohort details. Data are presented as median (interquartile range) where appropriate**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Patients, N</strong></td>
<td>149</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>64 (54-72)</td>
</tr>
<tr>
<td><strong>Gender (% Male)</strong></td>
<td>66.7</td>
</tr>
<tr>
<td><strong>Length of ICU Stay (days)</strong></td>
<td>8.4 (3.5-16.0)</td>
</tr>
<tr>
<td><strong>Operative/Non-operative</strong></td>
<td>49/100</td>
</tr>
<tr>
<td><strong>APACHE II score</strong></td>
<td>21.0 (15.0 – 25.0)</td>
</tr>
<tr>
<td><strong>Length of glycaemic control (hours)</strong></td>
<td>73.4 (43.2-135.7)</td>
</tr>
<tr>
<td><strong>Cohort total glycaemic control (hours)</strong></td>
<td>17,610</td>
</tr>
</tbody>
</table>
RESULTS

Table 3 presents a comparison of the performance of STAR and STOMP on the STAR cohort. STAR clinical data is presented for comparison with virtual trials using bolus insulin. An infusion-based virtual trial of STAR is also presented for direct comparison with the infusion-based STOMP protocol. Both STAR and STOMP allow variable measurement intervals, based on BG history. The value in Table 4 represents the maximum allowed measurement interval. Clinically, STAR uses a 3-hour maximum interval.

Table 3: Clinical and virtual trial performance for STAR and STOMP on STAR Cohort. **STAR Bolus Clinical** shows actual clinical performance of STAR. The other columns indicate virtual trial performance on this same cohort using bolus or infused insulin, with maximum measurement intervals as indicated.

<table>
<thead>
<tr>
<th></th>
<th>STAR Bolus Clinical</th>
<th>STAR Bolus 3hr max meas.</th>
<th>STAR Infusion 3hr max meas.</th>
<th>STOMP Infusion 3hr max meas.</th>
<th>STOMP Infusion 4hr max meas.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of measurements per day</td>
<td>13.4</td>
<td>13.2</td>
<td>12.7</td>
<td>10.0</td>
<td>8.4</td>
</tr>
<tr>
<td>% time BG 4.4 - 7 mmol/L</td>
<td>61.0</td>
<td>75.4</td>
<td>73.2</td>
<td>76.8</td>
<td>73.5</td>
</tr>
<tr>
<td>% time BG 4.4 - 8.0 mmol/L</td>
<td>80.6</td>
<td>87.5</td>
<td>86.3</td>
<td>87.7</td>
<td>86.2</td>
</tr>
<tr>
<td>% time BG 8.0 - 10 mmol/L</td>
<td>12.6</td>
<td>7.9</td>
<td>8.6</td>
<td>7.7</td>
<td>8.2</td>
</tr>
<tr>
<td>% time BG &gt; 10 mmol/L</td>
<td>5.1</td>
<td>2.5</td>
<td>2.6</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>% time BG &lt; 4.4 mmol/L</td>
<td>1.7</td>
<td>2.7</td>
<td>3.2</td>
<td>1.9</td>
<td>2.8</td>
</tr>
<tr>
<td>% time BG &lt; 2.2 mmol/L</td>
<td>0.006</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of Patients BG &lt; 2.2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Median Goal Feed [IQR] (%)</td>
<td>75.9 [36.4 - 99.4]</td>
<td>90.2 [30.7-100.0]</td>
<td>84.8 [30.5- 100.0]</td>
<td>80.0 [65.0 - 95.0]</td>
<td>80.0 [65.0 - 95.0]</td>
</tr>
<tr>
<td>Median insulin rate [IQR] (U/hr)</td>
<td>2.5 [1.0 - 4.0]</td>
<td>3.0 [1.5 - 5.5]</td>
<td>4.0 [1.5 - 6.0]</td>
<td>4.0 [2.0 - 5.5]</td>
<td>4.0 [2.0 - 5.5]</td>
</tr>
</tbody>
</table>
A randomly selected patient was chosen to show how the different controllers respond to fluctuating insulin sensitivity. Figure 4 shows this example and also shows how the total cost calculated by STOMP varies during a simulation. While overall the BG is largely similar, insulin interventions vary significantly, illustrating the differences between approaches.

Figure 4: Example showing the performance of STOMP and STAR (Bolus) on a patient. The target glycemic band (4.4-8.0 mmol/L) is shaded green in the top panel.
DISCUSSION

From the results it can be seen that the STOMP controller worked very well giving glycaemic results approximately equal to that of both STAR controllers. Both STAR and STOMP protocols demonstrated good performance. The percentage of time BG in band was relatively high, while also ensuring safety from hypoglycaemia (< 0.1% time below 2.2 mmol/L). Numbers of patients with BG < 2.2 mmol/L (4-5 patients) were lower than any reported study, where rates of 10-20% by patient have been reported [17, 45, 46]. Because STOMP optimises treatment and performance over a 6 hour window, it able to maintain excellent control performance with 4 hour measurement intervals. Hence, the average number of measurements per day and thus nurse GC workload are reduced by approximately 35%.

Figure 4 shows the total penalty cost for the chosen intervention in the bottom panel. STOMP can be seen to always be trying to minimize the intervention cost, and thus target an optimum balance of outcome-BG and intervention for the patient, based on the pre-selected weights and penalty functions. When the insulin sensitivity decreases, insulin becomes less effective and, as a consequence, larger interventions must be used to obtain the desired BG. This behaviour and choice causes a general increase in cost for all the potential interventions at that hour.

The method of using a combination of cost functions, long prediction horizon, stochastic prediction of variability, and a mathematical model to implement a glycaemic control protocol has proven to be very effective. This approach formalizes the STAR control algorithm, making it easier to optimize for different clinical requirements. Using cost functions to select the response of the controller means that changes in the response can be easily adapted to put a higher priority on any of the desired performance metrics, as clinically specified. Thus, clinical staff could choose the weighting for each performance metric based on the specific patient’s condition or local practice, making the protocol more patient-specific or hospital-specific. For example, if a patient was hyperglycaemic, but the clinicians wanted to continue giving the patient a high amount of nutrition, the cost functions could be adapted so there was a larger penalty for having low nutrition, thus causing the controller to have to choose an intervention which had a higher nutrition input, while also trying to maintain a desirable BG.
Additional cost functions, based on other aspects of glycaemic control, could be easily added to STOMP. This is a level of flexibility that is not currently available with the STAR protocol. For example, adding a cost function to reduce large changes in insulin interventions would make the controller less likely to over respond and suit a more cautious clinical practice. The MPC cost functions enabled this added flexibility in implementing a stochastic-predictive controller. STOMP is then potentially better suited for a more diverse range of clinical practice cultures and cohort-specific approaches than STAR.

It is also important to note that this MPC form of STAR is different than eMPC [47]. Beyond a different model, it uses stochastic model forecasting [28, 29] to manage intra-patient metabolic variability and risk [24, 48]. In contrast, eMPC uses auto-regressive models that are more patient-specific but require more data to adapt to changes. As a result this MPC approach (as with STAR) can safely predict further ahead because it does not rely on external black- or grey-box machine learning models to manage intra-patient variability.

The results of these virtual trials suggests that STOMP is ready for a pilot clinical trial. A pilot trial will validate these results on real patients prior to full clinical implementation.
CONCLUSIONS

An MPC glycaemic control protocol was developed and optimised for an adult BG signal in the ICU using the virtual trial approach. The STOMP protocol was designed as a model-predictive evolution of STAR that permits controller response to be easily tuned to specific, clinically relevant performance metrics. It has the additional benefit of formalising the heuristic control algorithm of STAR, and being a much more generalisable approach. The results indicate STOMP retains the performance and safety of STAR with approximately 85% time in the 4.4-8.0 mmol/L glycaemic band and 0.06% time < 2.2 mmol/L, with 35% fewer BG measurements. This protocol is a promising start that enables the development a more easily customized GC alternatives within an existing, proven model-based GC framework.

ACKNOWLEDGEMENTS

We acknowledge the support of a UC Summer Scholarship for Kent Stewart and a UC Doctoral Scholarship for Liam Fisk.
REFERENCES


