

Prediction Validation of Two Glycaemic Control Models in Critical Care^{*}

Ulrike Pielmeier^{*} J. Geoffrey Chase^{**} Steen Andreassen^{*}
Birgitte Steinfeldt Nielsen^{***} Pernille Haure^{***}
Geoffrey M. Shaw^{****}

^{*} Center for Model-based Medical Decision Support, Aalborg University,
Fredrik-Bajers-Vej 7, E4-215, 9220 Aalborg, Denmark (Tel:
+45 9635-7458; e-mail: upiel@hst.aau.dk)

^{**} Center for Bioengineering, Canterbury University, Christchurch,
New Zealand

^{***} Neuro- and Trauma Intensive Care Unit, Aalborg Hospital, Aalborg,
Denmark

^{****} Department of Intensive Care Medicine, Christchurch Hospital,
Christchurch, New Zealand

Abstract:

Metabolic models can substantially improve control of hyperglycaemia in critically ill patients. Control efficacy depends on how accurately a model-based system is able to predict future blood glucose (BG) concentrations after a glycaemic control intervention. This research compares two metabolic models in terms of their predictive power. Predictions 30 minutes to 10 hour forward are made using the Glucosafe model (GS) and a clinically validated model (CC) from Christchurch in a retrospective study of 11 hyperglycemic patients, 6 from New Zealand and 5 from Denmark. Median and ranges of prediction errors are similar for predictions up to 360 minutes. Both models make better predictions on the Danish patients. At long prediction times of more than 5 hours, GS predictions tend to be more accurate in the cohort from New Zealand whereas the CC model tends to predict better in the cohort from Denmark. However, relative differences in root mean square (RMS) of prediction errors are not greater than 4-5% in both cohorts. For both models, outlying prediction errors are dominated by single patients, particularly type 1 diabetic patients. GS predicted BG values are generally higher compared to CC predicted values. As expected, the RMS prediction error increases with prediction interval for both models and cohorts. Results show the potential of both models for use in prospective clinical trials with 120-180 min sampling intervals. Predictive power is attributed to the type of cohort in terms of degree of illness and glycaemic stability as well as sensor type used.

Keywords: Model-based control; Physiological Models; Decision support systems; Prediction error methods.

1. INTRODUCTION

Hyperglycaemia occurs frequently among patients in critical care (Umpierrez et. al. [2002], Krinsley [2003]). Glycaemic control can save lives, reduce morbidity and reduce hospital stay (Van den Berghe et. al. [2001]). However, resources are often limited, prohibiting strict regulated protocols, in addition to which increased hypoglycaemia has been reported (Van den Berghe et. al. [2006], Brunkhorst et. al. [2008]).

Systems based on metabolic models account for patient specific dynamics and can provide better glycaemic control without additional clinical intervention (Chase et. al. [2007]). Efficacy depends on the accuracy of predictions of future blood glucose concentrations (BG). The length of time ahead for which accurate predictions can

^{*} This work was partially supported by a grant awarded by the Programme Commission on Nanoscience, Biotechnology and IT under the Danish Council for Strategic Research.

be made will affect the clinical burden, which can limit performance (Shulman et. al. [2007], Chase et. al. [2006]).

This research compares two metabolic models in terms of their predictive power. Glucosafe (GS) is a new composite model that encompasses previous work on insulin appearance and diffusion (VanCauter et. al. [1992], Lotz [2007]), a glucose transporter model (Arleth et. al. [2000]), and new research on insulin binding and the absorption of carbohydrates in critically ill patients. It's compared against a clinically validated model (CC), that has been demonstrated to provide good glycaemic control in critical care with blood samples taken in one to two hour intervals (Shaw et. al. [2006]).

Two-hourly blood screens require more time and resources for some intensive care units that cannot be routinely afforded. The focus of this comparison is therefore on prediction times that might routinely exceed two hours.

Table 1. PATIENT CHARACTERISTICS AND BLOOD GLUCOSE CONTROL

Characteristic	Aalborg Patients					Christchurch Patients					
	A1	A2	A3	A4	A5	87	130	229	519	554	847
PATIENT CHARACTERISTICS											
Sex	F	F	F	M	M	F	M	F	M	F	F
Age	n.a.	58	75	60	46	62	21	73	69	20	67
Admission *	SAH	SAH	trauma	SDH	SAH	med.	trauma	card.	surg.	med.	med.
GCS	3	8	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
ApacheII	n.a.	n.a.	n.a.	n.a.	n.a.	26	11	15	29	26	17
Diabetes	—	—	—	2	—	—	1	—	2	1	—
BG upon admission (mmol/l)	6.2	5.9	10.8	11.6	7.1	10.1	14.1	15	11.5	10	12.7
BLOOD GLUCOSE CONTROL											
Hours	265	251	317	46	80	154	33	241	294	54	37
Blood Glucose (mmol/l)											
— mean	7.4	7.0	7.5	6.9	7.6	9.1	11.3	7.6	6.3	7.1	7.5
— max	11.9	9.0	11.2	11.6	9.7	13.9	16.6	15.0	11.5	11.8	12.7
— min	4.5	5.4	3.5	3.9	6.3	3.3	4.2	2.5	3.3	3.1	3.9
Sampling interval **	3h19	5h20	3h31	2h25	3h49	3h12	1h04	3h05	3h35	1h21	3h05
proportion in band ***	.44	.55	.32	.56	.16	.21	.12	.43	.65	.34	.52

* SAH: subarachnoid hemorrhage, SDH: subdural hematoma, card.: cardiologic, surg.: surgical, med.: other medical

** mean blood glucose sampling interval in (hours:minutes)

*** proportional time spent in 4–7 mmol/l band, calculated by linear interpolation

2. METHODS

Data from 11 patients in two cohorts were retrospectively gathered. The first cohort were six patients from a Christchurch, New Zealand (NZ) data pool that had been previously used for data fitting and validation of the CC model (Hann et. al. [2005]). They were chosen for a data density of ≤ 3.5 hours. Three patients were diabetics; this patient subgroup was also over-represented in the original data pool. Recordings of five consecutively observed hyperglycaemic patients from the Neuro- and Trauma Intensive Care Unit at Aalborg Hospital, Denmark (DK) constituted the second cohort. The first part of Table 1 lists the patient characteristics at admission.

Glycaemic control in both cohorts was conducted according to local department policies. In the DK group, blood samples were taken via a prior inserted (arterial) cannula and blood glucose concentrations determined by an ABL700 blood gas analyzer (Radiometer A/S). Arterial cannula samples with GlucoCard glucometers (Arcroy Inc., Japan) were used for the NZ group. These latter sensors have a larger measurement error up to 10%. All patients were enterally fed aside from two Aalborg patients who received supplementary intravenous glucose infusions.

Figure 1 shows the cumulative measured blood glucose distributions for both cohorts. The second part of Table 1 summarizes the glycaemic control data per patient. The NZ cohort is more variable, possibly reflecting more severe illness in these patients.

Glucosafe is a four-compartment model that combines research by VanCauter et. al. [1992], Lotz [2007] on insulin secretion and the glucose transporter model by Arleth et. al. [2000]. The model assumes a saturating insulin effect at high interstitial insulin concentrations. The intestinal glucose absorption rate as described by Arleth et. al. [2000] was decreased by a factor of two

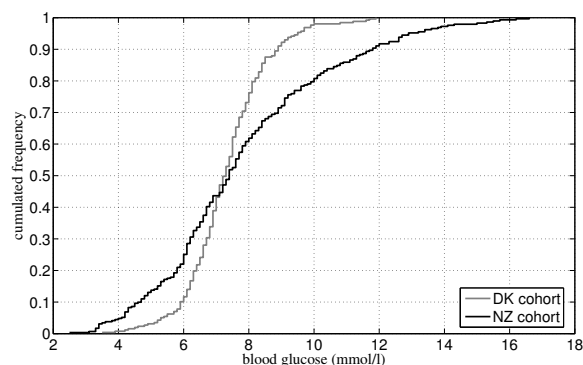


Fig. 1. Cumulative measured blood glucose concentrations of the Aalborg (DK) and the Christchurch (NZ) cohort.

to account for the often observed delay in gastric emptying in enterally fed critically ill subjects (Chapman et. al. [2007]). In the Glucosafe model, the endogenous insulin production rate and the insulin sensitivity are two patient parameters with a priori values of 40 U/day for the endogenous production of insulin and 0.1625 for the insulin sensitivity. In this model, the insulin sensitivity is a factor for the reduced insulin-mediated glucose uptake in hyperglycaemic critically ill patients relative to normal subjects with an insulin sensitivity of 1.0. The parameters were identified by a function that minimized the error from fitting the simulated BG concentration curve to the last measured datapoint and the difference of the parameters to their a priori values. The identified patient parameters were used to predict future blood glucose concentrations.

In the CC model, a suppression of endogenous insulin production is assumed at typically ample administration of exogenous insulin. The patient parameter that determines predictions of blood glucose is the insulin sensitiv-

ity. The implementation was done as described in Wong et. al. [2005], with the following modifications: 1) Insulin sensitivity was determined as the average value from the parametric fits of the last two datapoints, rather than using the integral method; 2) To prevent patients who did not receive a continuous insulin infusion from having zero plasma insulin, a constant "basal" insulin concentration was added at 11 mU/l (Katz et. al. [1993]).

In both models, the glucose distribution volume was set to 14 liters and the plasma insulin volume was calculated from body surface area and gender according to the method proposed by VanCauter et. al. [1992].

Predictions were made per patient per model by moving forward along the blood glucose measurements. At each measured point, blood glucose concentrations were calculated every minute over the following ten hours. Matching pairs of predicted values and prospective measurements were recorded together with the prediction time interval and the identified patient parameters. Errors were calculated in percent as predicted minus measured value divided by measured value. Median and total ranges of the percent errors were analysed per model per cohort, grouped by measured blood glucose concentration. The groups were arranged with approximately equal numbers of predictions in each group. The root mean square (RMS) of prediction error was determined for hourly time intervals per cohort and model, and means and standard deviations of prediction errors were calculated for individual patients. Software tools for statistical analysis were SPSS 15.0 and Matlab 7.0.

3. RESULTS

3.1 Patient cohorts

Median BG concentrations were the same in the two patient cohorts (NZ: 7.4 mmol/l; DK: 7.3 mmol/l), but the interquartile ranges differed, showing a wider spread in BG concentrations among Christchurch patients (NZ: 6.0(25%), 9.2(75%); DK: 6.6(25%), 8.0(75%)), as is also evident in Figure 1.

The average blood sampling time interval (\pm standard deviation) was 154 (\pm 64) minutes in the NZ cohort and 221 (\pm 64) minutes in the DK cohort. The mean proportional time in the normoglycaemic band (4 – 7 mmol/l) was slightly higher in the Aalborg group (0.41) than in the Christchurch group (0.38). Pearson's correlation test showed that neither initial blood glucose at beginning of treatment nor mean blood sampling interval were correlated to the time spent in the normoglycaemic band.

3.2 Prediction results

1066 and 614 pairs of measured and predicted blood glucose concentrations were collected for the NZ and DK cohort, respectively.

Figure 2 displays percent prediction errors, grouped by measured blood glucose range, for all predictions of maximum three hours ahead. Predictions from both models for the DK cohort were more accurate than for the NZ cohort, resulting in smaller ranges and fewer outliers. GS

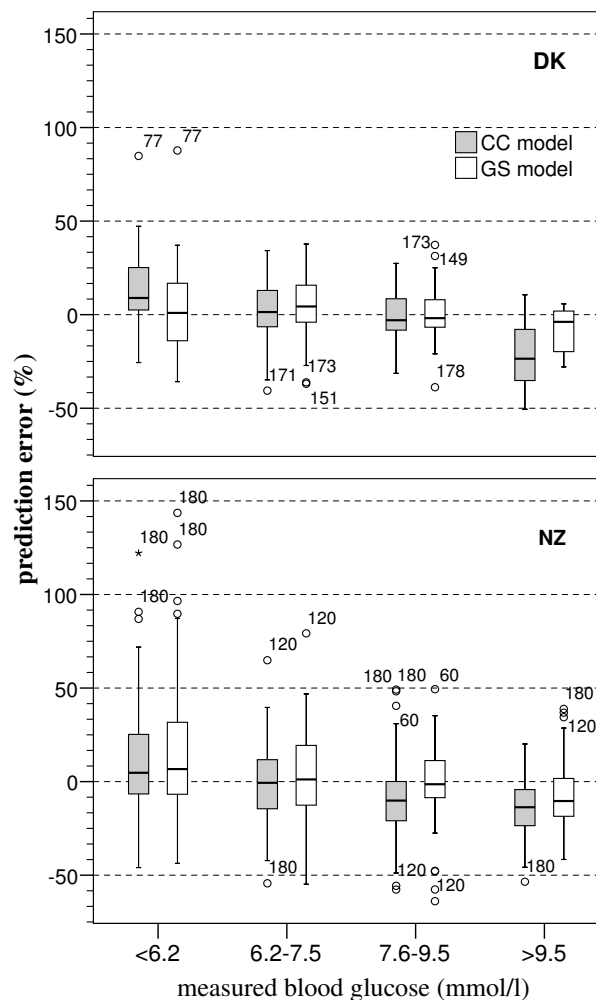


Fig. 2. Model prediction errors in percent, grouped by measured blood glucose range, for prediction times up to 180 minutes. Upper panel shows predictions for the Danish cohort, the lower panel for the cohort from New Zealand. Boxes include the range of scores falling into the middle 50% of the distribution (25th to 75th percentile) with medians as bold horizontal lines inside each box. Lines extend to the minimum and maximum scores in the distribution or to ± 1.5 times the interquartile range (IQR). Circles (asterisks) are outlying values more than 1.5 (3.0) times IQR from the first or the third quartile. Outliers are labeled with forward prediction times in minutes.

performed better in predicting the DK patients, whereas CC was more accurate predicting the NZ group.

As could be expected, error outliers were dominated by predictions of two to three hours ahead. A disproportionately high number of outliers resulted from predictions of patients 130, 554 and 229 (not displayed). Both models tended towards too low predictions for measurements that were very high and vice versa.

Errors for predictions of three to six hours ahead are shown in Figure 3. Prediction error ranges of the NZ cohort were again substantially larger than the DK cohort, though outliers were not primarily for longest prediction

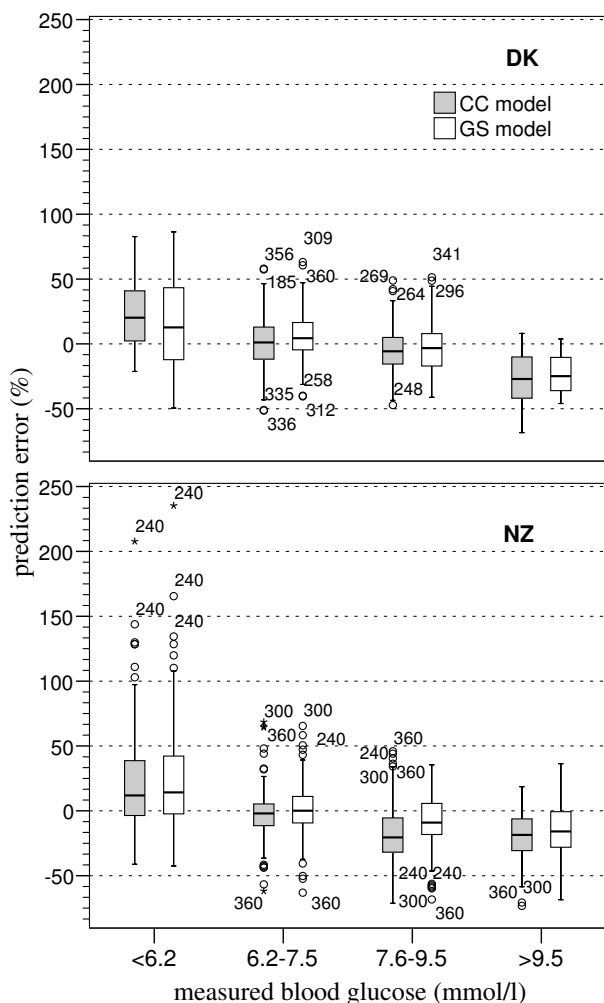


Fig. 3. Model prediction errors in percent, grouped by measured blood glucose concentration, for prediction times between 180–360 minutes. Upper panel shows the Danish cohort, the lower panel is the cohort from New Zealand. Labels of outliers denote forward prediction times in minutes.

times as might have been expected. The tendency towards underestimating high blood glucose measurements and too high predictions of low measurements recurred. Overall performance of the models was very similar.

Figures 4 and 5 give the RMS of the prediction errors, calculated from absolute deviations from measured values in mmol/l and from percent errors, respectively, in hourly intervals per model per patient group.

The predictive power of the two models matches for the NZ cohort over the first five hours, whereafter predictions by the GS model tend to be a little more accurate. For the NZ cohort the maximum difference in RMS of prediction error was 4.1% that was reached at prediction times of 420–480 minutes ahead.

Both models performed better on the cohort of the more glycaemically stable Aalborg patients. This result is expected given that both models tend to underpredict bigger changes. Performance was about equal up to 6 hours,

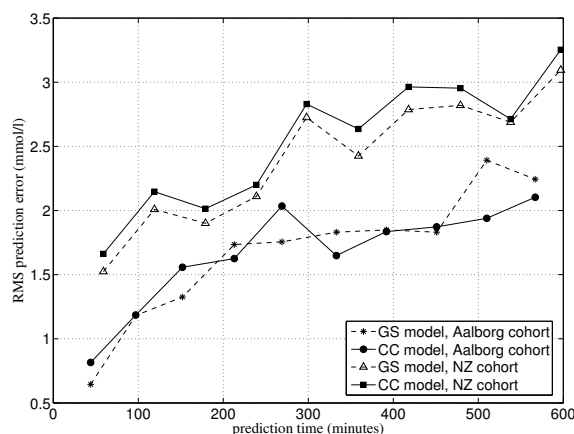


Fig. 4. RMS prediction error in mmol/l, calculated for hourly intervals. Smallest prediction time was 30 minutes, maximum prediction time 600 minutes. Solid lines: CC model. Dashed lines: GS model.

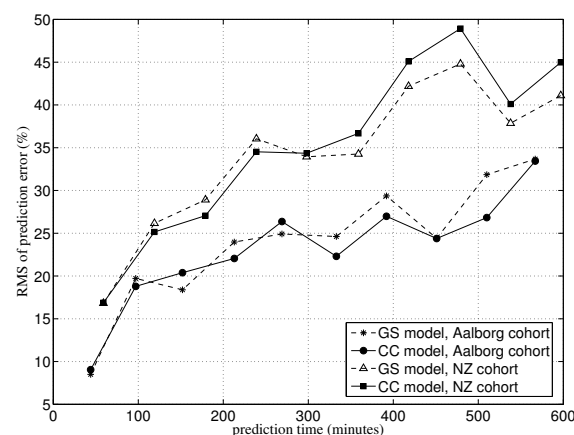


Fig. 5. RMS of the prediction error in percent. Solid lines: CC model. Dashed lines: GS model.

Patient	MPE GS	SD GS	MPE CC	SD CC
A1	1.4	29.3	-3	27.7
A2	6.1	16.2	.6	18.4
A3	5.0	23.8	3.7	19.3
A4	12.2	33.3	9.7	32.4
A5	2.6	15.7	-13.0	28.5
87	-5.4	18.6	-7.9	19.4
130	9.1	43.7	7.6	41.3
229	4.4	35.8	2.9	34.6
519	4.1	25.5	4.6	32.3
554	-5.3	38.3	-16.8	39.1
847	1.6	32.1	-1.7	37.9

Table 2. Means of percent prediction errors (MPE) and standard deviations (SD) for individual patients per applied model.

whereafter the CC model slightly outperformed the GS model, with a peak difference in RMS of prediction error of 5% (0.5 mmol/l) at prediction times of 480–540 minutes ahead.

Means of percent prediction errors (MPE) and standard deviations (SD) are shown for individual patients per

model in Table 2. In both models, prediction errors for the two type 1 diabetic patients from the NZ group (patients 130 and 554) were largest. In the GS model positive mean values predominate indicating a skewness of predictions towards too high or overpredicted values. Standard deviations in both models are comparable, and are smaller for the Aalborg patients.

On average, the prediction errors increased approximately linearly with progressing time. Figures 4 and 5 may be also used to determine a best prediction time for a cohort. For example, if a 25% error is the acceptable limit, then the NZ cohort should be predicted no more than 120 minutes ahead in either model. Similarly, for the same error level, the DK cohort can be “safely” predicted within 250 minutes for the CC model and 250–300 minutes for the GS model. This result is indicative of the difference in stability or level of critical illness between the cohorts with the NZ cohort being the more critically ill and likely more dynamic or unstable cohort.

4. DISCUSSION

This research compared two metabolic models with regard to their ability to make accurate predictions of blood glucose concentrations. Retrospective data of eleven patients from two different cohorts had been chosen for inclusion in the test. Short-term (minimum 30 minutes) and long-term (maximum 600 minutes) predictions were investigated and the difference in prediction errors was compared.

Across all analyses, both models made substantially better predictions on the DK group of patients. This result may be explained by the differences between the two cohorts. The cumulative measured blood glucose distributions in Figure 1 showed that the spread in the Christchurch group was wider than in the Aalborg group. Patients in the Aalborg group were more alike in terms of disease condition, whereas the Christchurch patients represented a broad cross-sectional mix including two type 1 diabetic patients. The wider distribution for the Christchurch group also indicates greater level of critical illness and thus greater dynamic behaviour, see Chase et. al. [2006].

Another, numerical cause may be the difference in glucose measurement error, with the DK cohort having much smaller assay errors. Numerically, smaller errors would allow more accurate parameter identification, and thus potentially better prediction (Chase et. al. [2006]). This assay error difference is clearly evident in the lower y intercept in Figures 4 and 5.

There was a pronounced tendency to more and larger outliers among NZ patients. Chase et. al. [2007] and [?] both indicated that the glucometers used in Christchurch patients could provide contaminated or significantly erroneous measurements. In their SPRINT study, such measurements with changes of over 3–4 mmol/l/hour were recorded approximately 1.5% of the time, but not evenly distributed. Hence, some outlying errors in the NZ cohort may be due to such measurements and are a fact of life in many intensive care units.

The overall error differences remained quite small. A slight advantage of the GS model in prediction errors could be made out for the NZ cohort, whereas the CC model

performed equally good or better for the DK cohort. However, the maximum differences in RMS of prediction errors amounted to only 4–5% in both cohorts at long prediction times of 420–540min.

The analysis of percent prediction errors grouped by measured blood glucose concentrations showed a tendency of both models towards underestimating high blood glucose values in contrast to overestimating values in the lower band. However, the majority of measurements was between 6 and 8 mmol/l, and probably many measurements above or below were preceded by measurements from within this range. Therefore, and because the models rely on information about past measurements, the skewed error distributions show a certain inert anticipation of sudden changes.

Results for individual patients indicated a more general bias in GS model predicted values, which tended to be slightly too high. The reason for this small bias could be that the a priori population values for the insulin sensitivity and the endogenous insulin production rate were set too low. An increase of both values would raise the effect on simulated blood glucose, possibly yielding improved predictions. Therefore, further research should be done to revise the currently used values.

It was mentioned in the results section that for both models a majority of outlying prediction errors was dominated by single patients. Namely three patients from the Christchurch cohort, of which two were type 1 diabetic patients. Of all eleven patients, they had the shortest sampling intervals. It could have been expected that predictions on these patients would be even more accurate for the tighter control they received. However, this was not the case. The authors showed in a previous study (Pielmeier et. al. [2008]) how admission type or diabetic status affect glycaemic predictability. In this research, the majority of outliers was generated by three-hour-forward predictions. Therefore doubts can be raised whether the models in their current shape sufficiently capture such highly fluctuating patients in such medium to long-term predictions. The results from this research thus encourage developing improvements to the models for this particular patient subgroup.

5. CONCLUSION

Overall, the results show the following salient facts:

- (1) Both models predict with similar accuracy.
- (2) The CC model could be used for longer predictions than it is currently.
- (3) The GS model has significant prediction capacity and potential for safe use in a large scale clinical trial.
- (4) The predictive power of the two tested models is restricted by the cohort type. A definition of cohorts could improve the integration of model-based advice into clinical routine.

Finally, the approach presented in this work is general enough to evaluate any similar metabolic model and may

be used to benchmark such central models and prediction times.

REFERENCES

- T. Arleth, S. Andreassen, M.O. Federici, and M.M. Benedetti. A model of the endogenous glucose balance incorporating the characteristics of glucose transporters. *Computer Methods and Programs in Biomedicine*, 62: 219–234, 2000.
- F. Brunkhorst, C. Engel, F. Bloos, A. Meier-Hellmann, M. Ragaller, N. Weiler, O. Moerer, M. Grndling, M. Oppert, S. Grond, D. Olthoff, U. Jaschinski, S. John, R. Rossaint, T. Welte, M. Schaefer, P. Kern, E. Kuhnt, M. Kiehnopf, C. Hartog, C. Natanson, M. Loeffler, K. Reinhart. Intensive Insulin Therapy and Pentastarch Resuscitation in Severe Sepsis. *N Engl J Med*, 358(2): 125–139, January 2008.
- M.J. Chapman, N.Q. Nguyen, and R.J. Fraser. Gastrointestinal motility and prokinetics in the critically ill. *Curr Opin Crit Care*, 13: 187–194, 2007.
- J.G. Chase, C.E. Hann, J. Wong, J. Lin, T. Lotz, A. LeCompte, and T. Lonergan. Overview of glycemic control in critical care: Relating performance and clinical results. *J Diabetes Sci Technol*, 1(1): 82–91, January 2007.
- J.G. Chase, C.E. Hann, M. Jackson, J. Lin, T. Lotz, and G.M. Shaw. Integral-based filtering of continuous glucose sensor measurements for glycaemic control in critical care. *J. of Computer Methods and Programs in Biomedicine*, 82(3): 238–247, 2006.
- J.G. Chase, G.M. Shaw, X.W. Wong, T. Lotz, J. Lin, and C.E. Hann. Model-based glycaemic control in critical care – A review of the state of the possible. *Biomedical Signal Processing and Control*, 1: 3–21, 2006.
- C.E. Hann, J.G. Chase, J. Lin, T. Lotz, C.V. Doran, and G.M. Shaw. Integral-based parameter identification for long-term dynamic verification of a glucose-insulin system model. *J. of Computer Methods and Programs in Biomedicine*, 77(3): 259–270, 2005.
- H. Katz, P. Butler, M. Homan, A. Zerman, A. Caumo, C. Cobelli, and R. Rizza. Hepatic and extrahepatic insulin action in humans: measurement in the absence of non-steady-state error. *Am J Physiol*, 264: 561–566, 1993.
- J.S. Krinsley. Association between hyperglycemia and increased hospital mortality in a heterogeneous population of critically ill patients. *Mayo Clin Proc*, 78(12): 1471–1478, 2003.
- T. Lotz. High Resolution Clinical Model-Based Assessment of Insulin Sensitivity. *PhD thesis*, Dept of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand, 2007.
- K.C. McCowen, A. Malhotra, and B.R. Bistrain. Stress-induced hyperglycemia. *Crit Care Clin*, 17(1): 107–124, 2001.
- U. Pielmeier, S. Andreassen, and J.G. Chase. When to measure blood glucose - Cohort-specific glycaemic control. *6th IASTED International Conference on Biomedical Engineering*, Innsbruck, Austria, February 2008.
- G.M. Shaw, J.G. Chase, X.W. Wong, J. Lin, T. Lotz, A. LeCompte, T. Lonergan, M. Willacy, and C.E. Hann. Rethinking glycaemic control in critical illness - from concept to clinical practice change. *Critical Care and Resuscitation*, 8(2): 90–99, June 2006.
- R. Shulman, S.J. Finney, C. O'Sullivan, P.A. Glynne, and R. Greene. Tight glycaemic control: a prospective observational study of a computerised decision-supported intensive insulin therapy protocol. *Critical Care*, 11: R75, 2007.
- W.D. Smith, A.G. Winterstein, T. Johns, E. Rosenberg, and B.C. Sauer. Causes of hyperglycemia and hypoglycemia in adult inpatients. *Am J Health Syst Pharm*, 62: 714–719, 2005.
- G.E. Umperiez, S.D. Isaacs, N. Bazargan, Y. Xiangdong, L.M. Thaler, and A.E. Kitabchi. Hyperglycemia: An independent marker of in-hospital mortality in patients with undiagnosed diabetes. *Journal of Clinical Endocrinology and Metabolism*, 87(3): 978–982, 2002.
- X.W. Wong, J.G. Chase, G.M. Shaw, C.E. Hann, T. Lotz, J. Lin, I. Singh-Levett, L.J. Hollingsworth, O.S.W. Wong, and S. Andreassen. Model predictive glycaemic regulation in critical illness using insulin and nutrition input: A pilot study. *Medical Engineering & Physics*, 27: 665–681, 2005.
- G. van den Berghe, P. Wouters et. al. Intensive insulin therapy in critically ill patients. *N Engl J Med*, 354(19): 1359–1367, 2001.
- G. van den Berghe, A. Wilmer et. al. Intensive insulin therapy in the medical ICU. *N Engl J Med*, 354(5): 449–461, 2006.
- E. van Cauter, F. Mestrez, J. Sturis, and S. Polonsky. Estimation of insulin secretion rates from C-Peptide levels. *DIABETES*, 41: 368–377, 1992.