International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:3, No:5, 2009

Extrapolation of Clinical Data from an Oral Glucose Tolerance Test Using a Support Vector Machine

Jianyin Lu, Masayoshi Seike*, Wei Liu, Peihong Wu, Lihua Wang, Yihua Wu,
Yasuhiro Naito, Hiromu Nakajima, and Yasuhiro Kouchi

Abstract—To extract the important physiological factors related to diabetes from an oral glucose tolerance test (OGTT) by mathematical modeling, highly informative but convenient protocols are required. Current models require a large number of samples and extended period of testing, which is not practical for daily use. The purpose of this study is to make model assessments possible even from a reduced number of samples taken over a relatively short period. For this purpose, test values were extrapolated using a support vector machine. A good correlation was found between reference and extrapolated values in evaluated 741 OGTTs. This result indicates that a reduction in the number of clinical test is possible through a computational approach.

Keywords—SVM regression, OGTT, diabetes, mathematical model

I. Introduction

A Reduction in the number of required blood samples and the duration of clinical tests, while maintaining the quality of the information needed is beneficial to both patients and medical staff. When patients are given tolerance tests that require time-dependent multiple blood samplings, it would be an advantage to be able to substitute estimated values for some of the test values by using extrapolation and interpolation techniques.

In previous studies of diabetes, we adopted mathematical modeling in order to extract important physiological factors [1] and are now applying this approach to oral glucose tolerance tests (OGTTs) as a model input. The OGTT is a common clinical test used to examine glucose tolerance under exogenous glucose load. This test is conducted in the morning after overnight fasting. At first, a fasting blood sample is taken. Then, the subject receives glucose orally and blood samples are taken again. The sampling schedules depend on the purpose of the test. For instance, glucose concentration at 120 min is defined as a diagnostic criterion of diabetes mellitus, where oral glucose load was applied at time 0 [2]. Mathematical modeling is used to process the OGTT result to assess physiological functions, such as insulin secretion and action, by analyzing the dynamics of plasma components.

Reducing the number of blood samples needed and the duration of the tests is necessary to enhance the usability of mathematical modeling [3]. At present, well validated models in this field rely on 300-min OGTT protocols in which plasma glucose and insulin, or plasma glucose and C-peptide concentrations have been sampled 11 times [4], [5]. However, such a large number of samples and prolonged period of study are not feasible for daily practice or large population-based studies. Hence, a convenient but informative protocol is required as an alternative.

The purpose of this study was to extrapolate clinical data to reduce the number of samples and duration of the OGTT, and to enhance the usability of the mathematical modeling approach for diabetes care. For this purpose, glucose and insulin concentration at 180 min were extrapolated using a support vector machine (SVM). The OGTT data from 741 subjects in various stages of glucose tolerance were analyzed to evaluate the estimation accuracy.

Manuscript received January 29, 2009; revised April 27, 2009. Asterisk indicates a corresponding author.

J. Lu, M. Seike*, and Y. Kouchi are with Central Research Laboratories, Sysmex Corporation, 4-4-4 Takatsukadai, Nishi-ku, Kobe 651-2271, Japan (e-mail*: Seike.Masayoshi@sysmex.co.jp).

W. Liu, P. Wu, L. Wang, and Y. Wu are with Department of Medicine and Endocrinology, Ren Ji Hospital, Shanhai Jiao Tong University School of Medicine, 1630 Dongfang Road Shanghai 200127, China.

Y. Naito is with Faculty of Environment and Information Studies, Keio University, 5322 Endo, Fujisawa 252-8520, Japan, with Institute for Advanced Biosciences, Keio University, 14-1 Baba-cho, Tsuruoka 997-0035, Japan, and with Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa 252-8520, Japan.

H. Nakajima is with Osaka Medical Center for Cancer and Cardiovascular Diseases, 1-3-3 Nakamichi, Higashinari-ku, Osaka 537-8511, Japan.

II. Materials and Methods

A. Subjects

Anonymous OGTT data from 741 subjects who had undergone an OGTT for clinical purposes at the Ren Ji Hospital in Shanghai, China were retrospectively analyzed. The data set consisted of 441 diabetes mellitus (DM) subjects, 167 impaired glucose tolerant subjects (IGT), regarded as pre-diabetes, and 133 normal glucose tolerant subjects (NGT). The characteristics of the subjects are shown in Table I.

TABLE I
Characteristics of subjects

	NGT	IGT	DM
Number	133	167	441
Basal Glucose (mg/dl)	90.2 ± 9.3	104.8 ± 12.4	143.3 ± 40.0
Basal Insulin (μU/ml)	8.2 ± 3.5	11.5 ± 9.6	13.5 ± 10.6
Age (years)	45.8 ± 13.8	55.5 ± 12.0	57.2 ± 11.9
BMI (kg/m^2)	23.6 ± 3.5	25.1 ± 3.4	24.9 ± 3.5

mean ± S.D.

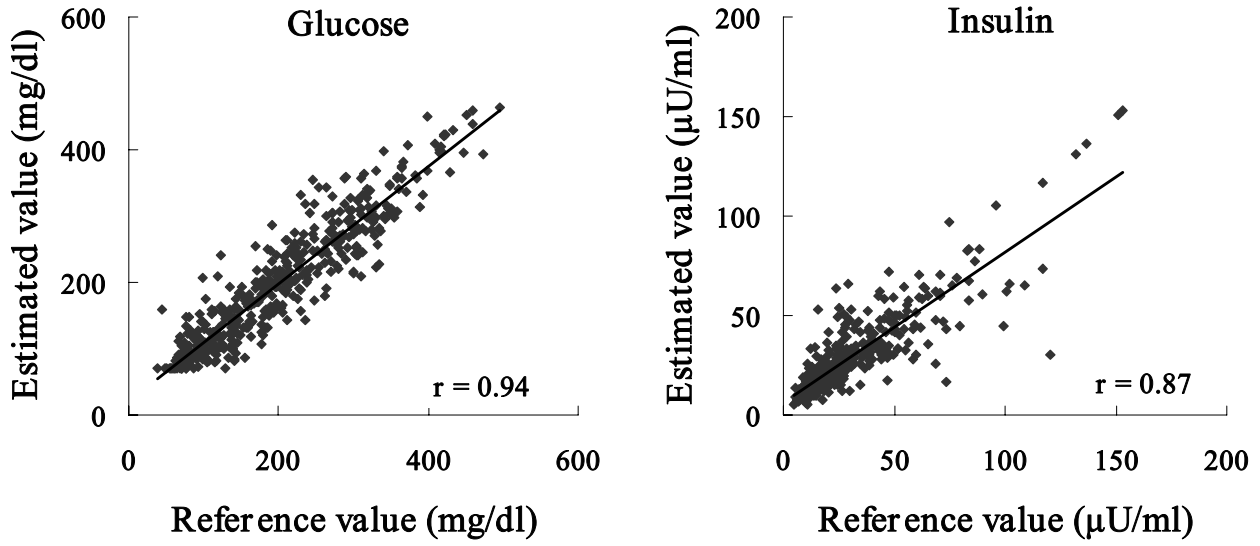

Fig. 1. Correlation between reference and estimated values at the 180 min point of the OGTT

B. OGTTs

The key regulator of glucose level is insulin. For this reason, glucose and insulin, or glucose and C-peptide concentrations are often selected for the model assessment of diabetes. In this study, plasma glucose and insulin concentrations were used. The oral glucose dose was 75 g and oral glucose load was applied at time 0. Plasma samples were collected at 0, 30, 60, 120, and 180 min.

C. Extrapolation

An SVM regression [6] (see Appendix) was applied to extrapolate glucose and insulin concentrations at 180 min from those sample values at 0, 30, 60, 120, and 180 min. Note that both glucose and insulin values were used for each glucose and insulin extrapolation, i.e.,

$$\begin{aligned} x &= (G_0, G_{30}, G_{60}, G_{120}, I_0, I_{30}, I_{60}, I_{120}) \\ y &= G_{180} \; or \; I_{180} \end{aligned}$$

where x represents the input, y represents the output, and G_t and I_t are glucose and insulin concentrations at time t, respectively. The training data set of 741 OGTTs was divided into 10 groups and the precision of the SVM regression was evaluated using 10-fold cross validation.

It is known that G_{180} rarely exceeds G_{120} by a large amount and the OGTT aims to keep the glucose level range over 70 mg/dl to avoid hypoglycemia. Accordingly, final extrapolated values were adjusted by introducing the following two conditions:

$$\begin{aligned} G_{180svm} &= G_{120} \qquad if \; G_{180svm} > G_{120} \\ G_{180svm} &= 70 \qquad if \; G_{180svm} < 70 \end{aligned}$$

where G_{180svm} represents the extrapolated glucose concentration at 180 min by SVM. The rationale for this consideration is also supported by the present data.

III. Results and Discussion

The estimated values for both glucose and insulin showed a significant correlation to the reference values (Fig.1). The average errors of extrapolation for glucose and insulin were 15.8% and 24.4%, respectively. Measurement accuracy in self-monitoring using blood glucose meters helps us interpret these results. These meters do not show exactly the same value as that measured in a clinical testing laboratory but are, nevertheless, informative. The International Organization for Standardization requires self-monitoring blood glucose meters to have an error rate of less than $\pm 20\%$ of the reference value in 95% of the measurements of glucose concentrations $\geq$ 75 mg/dl [7]. According to the criteria, the estimation of glucose using the SVM regression has sufficient clinical accuracy. There are no self-monitoring insulin meters, but the estimation of insulin would be useful.

The key finding of this study is the accuracy of extrapolation values at 180 min. As previously mentioned, the present validated models require 300-min protocols [4], [5]. However, in daily practice, the sampling schedule most often used to examine metabolic dynamics, apart from diagnosing diabetes, is 0, 30, 60, and 120 min. Accordingly, it is important to extrapolate values to 180, 240, and 300 min. Glucose and insulin concentrations during the test do not change biphasically within a period of 30 min and they hardly change biphasically within a period of 60 min after the start of the test. Thus, when values at 0, 30, 60, 120, 180, 240, and 300 min are obtained, it is not difficult to interpolate other data points. The values at 180 min have large variation especially among diabetic patients (Fig. 2 and Fig. 3) and values at 240 min have less individual variation than the values at 180 min. One of the reasons is that the rate of appearance of exogenous glucose in plasma varies greatly at 180 min [8]. Another reason is that insulin-dependent glucose uptake contributes to glucose regulation at 180 min, where the glucose level

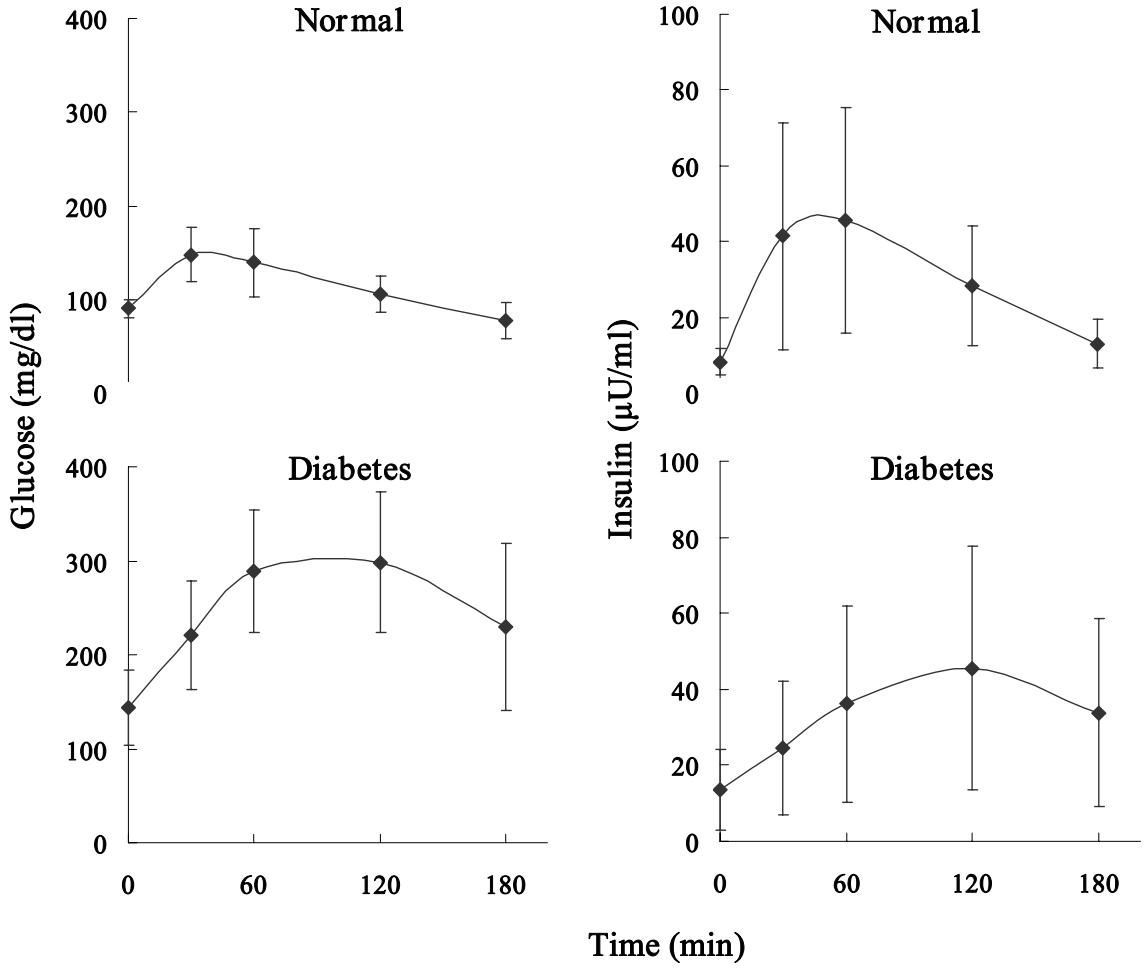

Fig. 2. Plasma glucose and insulin during the OGTT in normal glucose tolerant cases and diabetic patients (Mean ± S.D.)

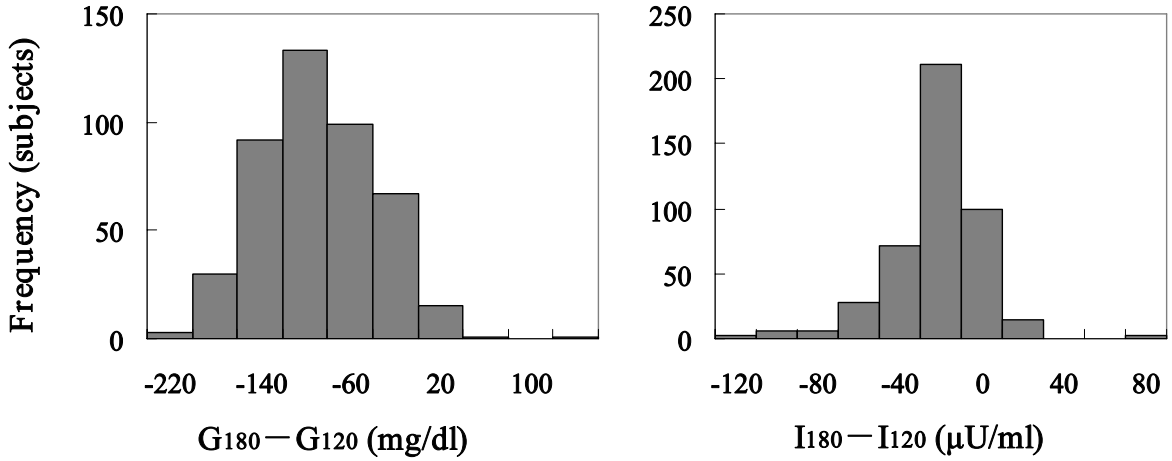

Fig. 3. Histogram of change of test values from 120 min to 180 min during the OGTT among diabetic patients

does not return to the basal level and insulin secretion is still stimulated. In contrast, both the appearance rate of exogenous glucose and insulin-dependent peripheral glucose uptake are not pronounced after 240 min. Values at 300 min return to the basal level in most cases. Therefore, if values at 180 min can be reasonably extrapolated from values within 120 min, it means that those models can be used even from values at 0, 30, 60, and 120 min. This 120-min 4-sample protocol is expected to facilitate the model assessment of diabetes in daily practice.

The distinguishing feature of the proposed procedure for the reduction of samples is that the original information of clinical data is maintained by the extrapolation. It is reported for non-diabetic subjects that the reproducibility of the model parameters was good between the full sampling (300-min, 11-sample) and reduced sampling (120-min, 7-sample) protocols [3] but it might not be accurate for diabetic patients as well as non-diabetic cases since it loses the informative variation at 180 min.

IV. Conclusion

Glucose and insulin concentrations at 180 min during the OGTT were extrapolated using a support vector machine. The OGTT data from 741 subjects in various stages of glucose tolerance were analyzed to evaluate the estimation accuracy. As a result, estimated values for both glucose and insulin showed good correlation to the reference values. The results showed that the 120-min 4-sample protocol combined with SVM could provide a reasonable basis to enhance the use of the mathematical modeling approach to diabetes care. This result also indicates that a reduction in clinical testing is possible by using the computational approach.

Appendix

For the SVM regression, a non-linear function is learned by a linear learning machine in a kernel-induced feature space [6]. A brief overview of this method is as follows.

Let x $\in \mathbb{R}^d$ and y $\in \mathbb{R}^1$ where $\mathbb{R}^k$ represents real space of k dimension. Given training data (x_1, y_1), ... , (x_n, y_n), the linear regression in some a priori chosen Hilbert space is given by

$$y = f(x, w) = \sum_{i=1}^{n} w_i \phi_i(x) + b \quad w = (w_1, ..., w_N, ...) \in \mathbb{R}^n$$

where $\phi_i(x)$ is the mapped vector of x in the chosen space and b represents a shift parameter.

Then, the optimization problem turns to finding a function that minimizes the following function, R(w)

$$R(w) = \frac{1}{n} \sum_{i=1}^{n} w_i |y_i - f(x_i, w)|_\varepsilon + \gamma < w, w >,$$

where

$$|y_i - f(x_i, w)|_\varepsilon = \begin{cases} 0 & if\ |y_i - f(x_i, w)| < \varepsilon \\ |y_i - f(x_i, w)| + \varepsilon & otherwise \end{cases}$$

$< w, w >$ is the inner product of two vectors, γ is some constant, and ε represents the tolerance criteria for estimation error.

The solution is given by the following equation

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) < \phi_i(x), \phi_i(x) > + b$$

where α_i, α_i^* are Lagrange multipliers, $\alpha_i, \alpha_i^* \geq 0$ with $\alpha_i \alpha_i^* = 0$, and $< \phi_i(x), \phi_i(x) >$ is the inner product of two elements of Hillbert space. Here, a Gaussian radial-basic-function kernel was used and it is described by the following equation:

$$< \phi_i(x), \phi_i(x) >= exp(-\frac{< x - x', x - x' >}{2\delta^2})$$

where x'$\in \mathbb{R}^d$ and δ is a constant.

In the present study, related dimensionless constants of δ and γ were set as 1.25 and 0.80 respectively.

Acknowledgment

We express special thanks to Mr. Kaoru Asano, Sysmex Corporation, for his helpful advice and encouragement.

References

[1] Y. Naito, H. Ohno, A. Sano, H. Nakajima, and M. Tomita, "Construction of a simulation model of diabetes for pathophysiological analysis using E-Cell System," *Genome Informatics*, vol. 13, pp. 478-479, Dec. 2002.

[2] K .G. M. M. Alberti and P. Z. Zimmet, "Definition, diagnosis and classification of diabetes mellitus and its complications. part. 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation," *Diabet Med*, vol. 15, no. 7, pp. 539-553, Jul. 1998.

[3] C. Dalla Man, M. Campioni, K. S. Polonsky, R. Basu, R. A. Rizza, G. Toffolo, and C. Cobelli, "Two-hour seven-sample oral glucose tolerance test and meal protocol: minimal model assessment of $beta$-cell responsivity and insulin sensitivity in nondiabetic individuals," *Diabetes*, vol. 54, no. 11, pp. 3265-3273, Nov. 2005.

[4] C. Dalla Man, A. Caumo, and C. Cobelli, "The oral glucose minimal model: estimation of insulin sensitivity from a meal test," *IEEE Trans Biomed Eng*, vol. 49, no.5, pp. 419-429, May. 2002.

[5] G. Toffolo, E. Breda, M. K. Cavaghan, D. A. Ehrmann, K. S. Polonsky, and C. Cobelli, "Quantitative indexes of beta-cell function during graded up&down glucose infusion from C-peptide minimal models," *Am J Physiol Endocrinol Metab*, vol. 280, no. 1, pp. E2-E10, Jan. 2001.

[6] V. Vapnik, S. Golowich, and A. J. Smola, "A support vector method for function approximation, regression estimation, and signal processing," in *Advances in Neural Information Processing Systems: Proceedings of the 1996 conference on Neural Information Processing Systems*, vol. 9, M. C. Mozer, M. I. Jordan, and T. Petsche, Ed. Cambridge: MIT Press, 1997, pp. 281-287.

[7] ISO15197, "In vitro diagnostic test systems: Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus," May. 2003.

[8] C. Dalla Man, M. Camilleri, and C. Cobelli, "A system model of oral glucose absorption: validation on gold standard data," *IEEE Trans Biomed Eng*, vol. 53, no. 12 Pt1, pp. 2472-2478, Dec. 2006.