# A Simple Prediction Rule and a Neural Network Model to Predict Pancreatic Beta-Cell Reserve in Young Adults with Diabetes Mellitus

SRIURAI THAMPRAJAMCHIT, M.D.*, SIRINATE KRITTIYAWONG, M.D.*, PONGAMORN BUNNAG, M.D.*, GOBCHAI PUAVILAI, M.D.*
BOONSONG ONGPHIPHADHANAKUL, M.D.*, SUWANNEE CHANPRASERTYOTHIN, M.Sc.*, RAJATA RAJATANAVIN, M.D.*,

**Abstract**

In the present study we developed and assessed the performance of a simple prediction rule and a neural network model to predict beta-cell reserve in young adults with diabetes. Eighty three young adults with diabetes were included in the study. All were less than 40 years old and without apparent secondary causes of diabetes. The subjects were randomly allocated to 2 groups; group 1 (n = 59) for developing a prediction rule and training a neural network, group 2 (n = 24) for validation purpose. The prediction rule was developed by using stepwise logistic regression. Using stepwise logistic regression and modification of the derived equation, the patient would be insulin deficient if 3(waist circumference in cm) + 4(age at diagnosis) < 340 in the absence of previous diabetic ketoacidosis (DKA) or < 400 in the presence of previous DKA. When tested in the validation set, the prediction rule had positive and negative predictive values of 86.7 per cent and 77.8 per cent respectively with 83.3 per cent accuracy while the ANN model had a positive predictive value of 88.2 per cent and a negative predictive value of 100 per cent with 91.7 per cent accuracy. When testing the performance of the prediction rule and the ANN model compared to the assessment of 23 internists in a subgroup of 9 diabetics whose age at onset was less than 30 years and without a history of DKA, the ANN had the highest ability to predict beta-cell reserve (accuracy = 88.9), followed by the prediction rule (accuracy = 77.8%) and assessments by internists (accuracy = 60.9%).

We concluded that beta-cell reserve in young adults with diabetes mellitus could be predicted by a simple prediction rule or a neural network model. The prediction rule and the neural network model can be helpful clinically in patients with mixed clinical features of type 1 and type 2 diabetes.

**Key word :** Diabetes Mellitus, Beta-Cell Reserve, Beta-Cell Function, Prediction Rule, Neural Network

**THAMPRAJAMCHIT S, ONGPHIPHADHANAKUL B, KRITTIYAWONG S, et al**
**J Med Assoc Thai 2001; 84: 332-338**

* Department of Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand.

Diabetes mellitus (DM) is etiologically and clinically heterogeneous. The recently revised classification and diagnosis of DM is based on etiology and pathogenesis rather than solely on clinical features[1,2]. In clinical practice, however, limitations still exist in assessing beta-cell reserve in order to choose the most appropriate treatment and insulin injection may be introduced unnecessarily in some young diabetic adults. Although serum C-peptide levels may be used as a noninvasive mean in assessing pancreatic beta-cell reserve [3,4], the assay may not be readily available.

Clinical prediction rules based on clinical features have been derived to classify patients according to the risks of diseases[5]. Conventional statistical methods such as stepwise logistic regression, multiple linear regression and discriminant analyses are mostly used. These methods, although adequate in certain situations, can have reduced accuracy in non-linear systems. Moreover, the complexity of some equations makes it difficult to use in general clinical practice. Recently, artificial neural network (ANN) was introduced in clinical diagnosis and classification[6]. The ANN consists of a set of simple units that process information in parallel[7,8]. In some instances, ANN has been demonstrated to be more accurate than statistically derived prediction rules. It was the purpose of the present study to develop a simple prediction rule and a neural network model for predicting beta-cell reserve in young adults with diabetes and assess their accuracy.

## MATERIAL AND METHOD

### Patient

Subjects consisted of 83 young Thai adults with diabetes aged 14-40 years who were without apparent secondary causes of diabetes. The subjects were recruited from the diabetic outpatient clinics of Ramathibodi Hospital, Mahidol University and Theptarin General Hospital, Thailand from 1996 to 1997. The study protocol was approved by the Ethical Clearance Committee on Human Rights Related to Research Involving Human Subjects of the Faculty of Medicine, Ramathibodi Hospital, Mahidol University. Informed consent was obtained from each patient before investigation.

### Assays

Serum C-peptide level was measured by radioimmunoassay (Incstar, Stillwater, MN) after an overnight fast and at 6 minutes after an intravenous injection of 1 mg of glucagon. Serum was stored at -20°C for less than 6 months before the assay. The cutoff levels for basal and post-glucagon C-peptide were obtained from studying 22 young adults with DM. Based on clinical criteria, 10 of the subjects were classified as insulin-deficient and 12 subjects were insulin-sufficient. The clinical criteria for insulin deficiency included history of DKA, duration of diabetes of more than 2 years and insulin dependency. Clinical criteria for insulin sufficiency included the absence of DKA, absence of marked ketonuria during follow-up, good glycemic control while on oral agents and duration of diabetes of less than 5 years. The cut off points with 100 per cent sensitivity and 90 per cent specificity for poor beta-cell reserve were 0.36 nmol/L and 0.52 nmol/L for basal C-peptide and stimulated C-peptide, respectively. Subjects who had stimulated C-peptide above 0.52 nmol/L were classified as insulin-sufficient. The cutoff values used in previous studies include stimulated C-peptide level of 0.6 nmol/L[9,10], whereas, basal and stimulated C-peptide of 0.20 and 0.32 nmol/L were used in another study[4]. The difference in cutoff values may reflect the distinction of the assay systems.

### Method

Subjects were randomly allocated to 2 groups. The first group had 59 patients for developing a prediction rule and training an ANN. The second group consisted of 24 patients for validation. The training set initially was used to develop the prediction rule by stepwise logistic regression analysis. Transformation was performed on the derived equation to make it simpler for general use. The simplified rule was then tested in the validation set.

ANN was built and trained by using commercial software (Brainmaker, California Scientific Software, USA). The network consisted of 4 layers; 1 input layer with 10 nodes, 2 hidden layers with 10 nodes each and 1 output layer with 1 node. Each node in the input layer corresponded to each variable used in the training of the ANN. Clinical features entered into the ANN model included, gender, age at diagnosis, family history of DM, previous DKA, marked ketouria, body weight, height, BMI, waist circumference, hip circumference. Supervised training of ANN was done by presenting data from each subject in the training set to the ANN model until 100 per cent accuracy

**Table 1. Comparison of clinical characteristics and insulin reserve between the training set and the validation set.**

| Variables | Training set (n=59) | Validation set (n=24) | P-value |
|---|---|---|---|
| Age (year) | 31.4 ± 6.9 | 28.7 ± 7.1 | NS |
| Age at diagnosis (year) | 26.3 ± 8.3 | 23.8 ± 8.2 | NS |
| Sex (male:female) | 20:39 | 12:12 | NS |
| Duration of DM (year) | 5.2 ± 5.2 | 4.8 ± 5.1 | NS |
| BMI (kg/m$^2$) | 23.1 ± 4.1 | 23.9 ± 6.4 | NS |
| Waist (cm) | 77.7 ± 10.1 | 82.1 ± 16.7 | NS |
| W/H ratio | 0.83 ± 0.07 | 0.85 ± 0.08 | NS |
| Number of subjects with previous DKA | 16 | 9 | NS |
| Number of subjects with insulin deficiency | 31 | 15 | NS |

**Table 2. Comparison of clinical characteristics between insulin deficient group and insulin sufficient group in the training set.**

| Variables | Insulin deficient (n = 31) | Insulin sufficient (n = 28) | P-value |
|---|---|---|---|
| Age (year) | 29.2 ± 7.5 | 33.9 ± 5.4 | < 0.05 |
| Age at diagnosis (year) | 22.3 ± 7.8 | 30.8 ± 6.5 | < 0.05 |
| Sex (male:female) | 10:21 | 10:18 | NS |
| BMI (kg/m$^2$) | 21.5 ± 3.5 | 24.8 ± 4.1 | < 0.05 |
| Duration of DM (year) | 6.9 ± 5.8 | 3.1 ± 3.7 | < 0.05 |
| Waist (cm) | 72.5 ± 7.9 | 83.4 ± 9.2 | < 0.05 |
| W/H ratio | 0.80 ± 0.06 | 0.87 ± 0.05 | < 0.05 |
| Number of subjects with previous DKA | 14 | 2 | < 0.05 |

was achieved before testing in the validation set. Both methods were also tested in a subgroup of 9 young diabetics from the validation set whose age at onset was not more than 30 years, no history of DKA and the type of diabetes was uncertain based on clinical features alone. The results were then compared to the clinical assessment by 23 internists consisting of 1st and 2nd year internal medicine residents who had at least 4 years' experience in medical practice using a questionnaire consisting of the clinical data of these 9patients such as sex, age, body weight, presenting symptoms, history of DKA, treatment, family history of diabetes and ketonuria.

### Statistical analysis

Data was expressed as mean ± S.D. Differences in variables were determined by Student's *t*-test for continuous variables and chi-square test for categorical variables. P value less than 0.05 was considered to be statistically significant.

## RESULTS

There was no significant difference in clinical characteristics and beta-cell reserve between the training set and the validation set as shown in Table 1. As demonstrated in Table 2, it was found that age, age at diagnosis, body mass index (BMI), duration of diabetes, waist circumference, waist/hip (W/H) ratio and history of diabetic ketoacidosis (DKA) were significantly different between the insulin-deficient and the insulin-sufficient groups in the training set. From stepwise logistic regression analysis, previous DKA, lower waist circumference and lower age at diagnosis were significantly related to insulin deficiency (Table 3). After transformation as demonstrated in the Appendix, the following equations were obtained.

**Table 3. Factors related to insulin deficiency identified by logistic regression.**

| Variables | Odd Ratio | P value |
|---|---|---|
| Previous DKA | 23.6 | < 0.01 |
| Waist circumference (cm) | 1.2 | < 0.01 |
| Age at diagnosis (year) | 1.2 | < 0.01 |

**Table 4A. Prediction of beta-cell reserve: the simplified rule.**

| Prediction | Beta-cell reserve | |
|---|---|---|
| | Deficient | Sufficient |
| Deficient | 13 | 2 |
| Sufficient | 2 | 7 |

Accuracy = 83.3%
Positive predictive value = 86.7%
Negative predictive value = 77.8%

**Table 4B. Prediction of beta-cell reserve : ANN.**

| Prediction | Beta-cell reserve | |
|---|---|---|
| | Deficient | Sufficient |
| Deficient | 15 | 2 |
| Sufficient | 0 | 7 |

Accuracy = 91.7%
Positive predictive value = 88.2%
Negative predictive value = 100%

**The patient would be insulin-deficient if:**
3(waist circumference) + 4(age at diagnosis) < 340 in the absence of DKA or < 400 in the presence of DKA.

When tested in the validation set, the simplified prediction rule had 83.3 per cent accuracy with positive and negative predictive values of 86.7 per cent and 77.8 per cent, respectively (Table 4A) while the ANN model had 91.7 per cent accuracy, a positive predictive value of 88.2 per cent and a negative predictive value of 100 per cent (Table 4B). Although the ANN model appeared to have a higher positive predictive value, negative predictive value and accuracy compared to the simplified rule, these did not reach statistical significance. When testing the performance of the simplified rule and the ANN model compared to the assessments of 23 internists in a subgroup of 9 young diabetics whose age at onset was less than 30 years without a history of DKA, it was found that the ANN model had the highest ability to predict beta-cell reserve (8 correct predictions out of 9, accuracy = 88.9%), followed by the simplified rule (7 correct predictions out of 9, accuracy = 77.8%) and assessments by internists (126 correction predictions out of 207, accuracy = 60.9%).

## DISCUSSION

Correct classification of diabetic subjects at the time of diagnosis is sometimes difficult especially in young adults. Several reports have confirmed that C-peptide determinations are of value in the classification and choice of treatments[3,4]. In addition, HLA typing or measurement of islet cell autoantibodies may help to differentiate the type of DM[11]. These tests, however, are not readily available and thus are not as helpful as anticipated in clinical practice. Clinical parameters such as age at onset, body weight, ketosis and family history of diabetes have been used to discriminate between type 1 and type 2 diabetes but uncertainty still exists. In keeping with other previous studies[12,13], our study also found that younger age at onset, the presence of history of DKA and less waist circumference were predictors of insulin deficiency in young adults with diabetes. Waist circumference has been shown to be a better marker than W/H ratio for assessing abdominal fat [14] which is related to insulin resistance known to be pathophysiologically important in type 2 diabetes. One of the problems with the utilization of prediction rules in clinical practice is that the complexity of some of the prediction rules can prevent the rule from being used in the care of patients. Modification of the derived equation was thus performed in the present study. After simplification, it was found that the simplified rule had high accuracy, positive predictive value and negative predictive value. With its simplicity, more practicality in the assistance to estimate the status of beta-cell reserve can be expected. It is also of note that clinical assessment by 23 internists had low accuracy (60.7%) in the subgroup that had mixed features of type 1 and type 2. It follows that treatment based on clinical assessments may be

inappropriate in about half of the patients with similar features. In such cases, the simplified rule and the ANN model had more acceptable accuracy (77.8%, 88.9%). Therefore, when the status of beta-cell reserve is unclear from clinical assessment, the combined use of clinical characteristics and either the simplified rule or the ANN model will be helpful in distinguishing between type 1 and type 2 diabetes.

ANN has gained more utilization in clinical diagnosis and classification due to its superiority in terms of accuracy in certain situations[15-18]. The reason for the superiority may be due to the fact that most biological systems are nonlinear which can be better represented by ANN. In clinical practice, however, ANN may not be readily available. Moreover, there are certain factors which may influence the accuracy of the ANN model in classification problems. Apart from the training set being representative of the actual problem, the intrinsic relationship of the outcome of interest and its associated factors is also important. In situations where there are less than adequate causal or associative relationships between risk factors and the outcome of interest, the accuracy of the classification derived from the prediction models will not be good enough regardless of the computational models used. Although our present study found that the ANN model tended to have higher accuracy, positive predictive value and negative predictive value than the simplified rule, this did not reach statistical significance. This suggests that the nonlinearity which exists in the mapping of clinical variables to beta-cell reserve is modest and not much further improvement compared to linear methods can be obtained.

We concluded that beta-cell reserve in young adults with diabetes can be predicted by a simple prediction rule or a neural network model with reasonable predictive values. The prediction rule and the neural network model can be helpful clinically in patients with mixed clinical features of type 1 and type 2 diabetes.

(Received for publication on July 23, 1999)

## REFERENCES

1. The expert committee on the diagnosis and classification of diabetes mellitus: Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 1997; 20: 1183-97.
2. Alberti KG, Zimmet PZ. 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 1998; 15: 539-53.
3. Hother-Nielsen O, Faber O, Sorenson NS, Beck Neilson H. Classification of newly diagnosed diabetic patients as insulin-requiring or non-insulin-requiring based on clinical and biochemical variable. Diabetes Care 1988; 11: 531-7.
4. Gjessing HJ, Matzen LE, Faber OK, Froland A. Fasting plasma c-peptide, glucagon stimulated plasma c-peptide and urinary c-peptide in relation to clinical type of diabetes. Diabetologia 1989; 32: 309-11.
5. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Application and methodological standards. N Eng J Med 1985; 313: 793-9.
6. Katz WT, Snell JW, Merickel MB. Artificial neural networks. Methods Enzymol 1992; 210: 610-36.
7. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. The Lancet 1995; 346: 1075-9.
8. Penney W, Frost D. Neural networks in clinical medicine. Med Decis Making 1996; 16: 386-98.
9. Madsbad S, Krarup T, Mcnair P, et al. Practical clinical value of the c-peptide response to glucagon stimulation in the choice of treatment in diabetes mellitus. Acta Med Scand 1981; 210: 153-6.
10. Koskinen P, Viikari J, Irjala K, Kaihola HL, Seppala P. Plasma and urinary c-peptide in the classication of adult diabetics. Scad J Clin Lab Invest 1986; 46: 655-63.
11. Kaufman FR. Diabetes in children and adolescents: area of controversy. Med Clin North Am 1998; 82: 721-38.
12. Laakso M, Pyorola K. Age at onset and type of diabetes. Diabetes Care 1985; 8: 114- 7.
13. Melton LJ, Palumbo PJ, Chu CP. Incidence of diabetes by clinical type. Diabetes Care 1983; 6: 75-86.
14. Expert panel on the identification, evaluation and treatment of overweight and obesity in adults. Executive summary of the clinical guidelines on the identification, evaluation and treatment of

overweight and obesity in adults. Arch Intern Med 1998; 158: 1855-67.
15. Doyle HR, Dvorchik I, Mitchell S, et al. Predictive outcomes after liver transplantation. A connectionist approach. Ann Surg 1994; 219: 408-15.
16. Husain OA, Butler EB, Nayagam M, Mango L, Alonzo A. An analysis of the variation of human interpretation : Papnet a mini-challenge. Anal Cell Pathol 1994; 6: 157-63.
17. Maclin PS, Dempsey J. How to improve a neural network for early detection of hepatic cancer. Cancer Lett 1994; 77: 95-101.
18. Astion ML, Wener MH, Thomas RG, Hunder GG, Bloch DA. Application of neural networks to the classification of giant cell arteritis. Arthritis Rheum 1994; 37: 760-70.

## APPENDIX

Let A = age at diagnosis
W = waist circumference
H = history of DKA (2 = presence, 1 = absence)
From logistic regression model:
Probability of sufficient beta-cell reserve $= \frac{1}{1+e^{-z}}$

where $z = 0.21\,A + 0.15\,W + 3.2\,H - 23.1$. (1)
Cutoff probability of sufficient beta-cell reserve = 0.5 if $z = 0$.
Therefore the subject is insulin deficient if $z < 0$. (2)

From (1) and (2), beta-cell reserve is poor when
$0.21\,A + 0.15\,W + 3.2\,H - 23.1 < 0$
which is approximately, $\frac{4\,A + 3\,W + 62\,H}{20} - 23.1 < 0$
or $4\,A + 3\,W < 400$ in the presence of history of DKA,
$4\,A + 3\,W < 340$ in the absence of history of DKA.

# การทำนาย beta-cell reserve ในผู้ป่วยเบาหวานอายุน้อยด้วย prediction rule อย่างง่ายและ artificial neural network

*ศรีอุไร ธรรมประจำจิต, พ.บ.*, บุญส่ง องค์พิพัฒนกุล, พ.บ.*, สิริเนตร กฤติยาวงค์, พ.บ.*, สุวรรณี ชั้นประเสริฐโยธิน, วท.ม.*, พงศ์อมร บุนนาค, พ.บ.*, รัชตะ รัชตะนาวิน, พ.บ.*, กอบชัย พัววิไล, พ.บ.**

ผู้วิจัยได้พัฒนา prediction rule อย่างง่าย และ artificial neural network (ANN) สำหรับทำนาย beta-cell reserve ในผู้ป่วยเบาหวานอายุน้อยและทดสอบความแม่นยำในการศึกษานี้ ผู้เข้าร่วมการศึกษาประกอบด้วยผู้ป่วยเบาหวานไทยอายุน้อยกว่า 40 ปีจำนวน 83 รายที่ไม่มีสาเหตุอื่นแน่ชัด ทำการพัฒนา prediction rule และ ANN จากข้อมูลของผู้ป่วย 59 ราย และทำการทดสอบความแม่นยำในผู้ป่วยที่เหลือ 24 ราย ได้ prediction rule คือ ผู้ป่วยจะมี beta-cell reserve ไม่เพียงพอหาก 3(เส้นรอบวงเอวหน่วยเป็นซม.) + 4(อายุที่เริ่มเป็นเบาหวาน) มีค่าน้อยกว่า 340 ในผู้ที่ไม่เคยมีประวัติ DKA หรือ น้อยกว่า 400 ในผู้ที่เคยมีประวัติ DKA สูตรดังกล่าวมีความแม่นยำร้อยละ 83.3 ในขณะที่ ANN มีความแม่นยำร้อยละ 91.7 เมื่อทำการทดสอบเพิ่มเติมในผู้ป่วยเบาหวาน 9 รายที่เริ่มเป็นเมื่ออายุน้อยกว่า 30 ปี โดยไม่มีประวัติ DKA พบว่าการประเมินโดยแพทย์ 23 รายทำนาย beta-cell reserve ได้ถูกต้องร้อยละ 60.9 ในขณะที่ prediction rule ทำนายได้ถูกต้องร้อยละ 77.8 และ ANN ถูกต้องร้อยละ 88.9 สูตรอย่างง่ายและ ANN นี้น่าจะเป็นประโยชน์ในการช่วยประเมิน beta-cell reserve และเลือกการรักษาที่เหมาะสมสำหรับผู้ป่วยเบาหวานอายุน้อยที่มีลักษณะทางคลินิกผสมระหว่างเบาหวานชนิดที่ 1 และ 2

**คำสำคัญ** : เบาหวาน, Beta-Cell Reserve, การทำงานของ Beta-Cell, สูตรทำนาย, Neural Network

**ศรีอุไร ธรรมประจำจิต, บุญส่ง องค์พิพัฒนกุล, สิริเนตร กฤติยาวงค์, และคณะ จดหมายเหตุทางแพทย์ ฯ 2544; 84: 332-338**

* ภาควิชาอายุรศาสตร์, คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี, มหาวิทยาลัยมหิดล, กรุงเทพฯ 10400