A Probabilistic Model for Predicting Hypoglycemia in Type 2 Diabetes Mellitus

Background: To develop and validate a method for estimating hypoglycemia risk in stable, insulin-treated subjects with type 2 diabetes mellitus.

Methods: Subjects (n=195) monitored their blood glucose levels 4 times daily for 8 weeks. An 8-week mean blood glucose value (GLUMEAN) with standard deviation (GLUSD) was derived for each patient. Subjects were then randomly allocated to a derivation or validation set. For the derivation set, we developed a logistic function based on GLUMEAN and GLUSD to describe the 8-week risk of hypoglycemia (blood glucose \(< 60 \text{ mg/dL} [3.3 \text{ mmol/L}]\)). This function was used to assign a predicted probability of hypoglycemia to each subject in the validation set. Subjects were assigned to risk quartiles and followed up for up to 52 weeks.

Results: We evaluated 195 subjects, 95% of whom were men and 69% of whom were non-Hispanic white. For 72 derivation subjects, GLUMEAN and GLUSD were highly influential determinants of hypoglycemia during intensified monitoring. The 123 validation subjects were followed up for 39.7±7.1 weeks (mean±SD). The occurrence of long-term hypoglycemia differed significantly across risk quartiles (19.4%, 36.7%, 61.3%, and 77.4%, respectively; \(P<.001\)). Receiver operating characteristic curve analysis showed that the area for the probability function (0.746±0.046) was significantly higher than the area for hemoglobin A1c (0.549±0.052) because their 95% confidence intervals did not overlap. The function also identified subjects who developed long-term hypoglycemia at a rate exceeding the median frequency.

Conclusions: Self-monitoring of blood glucose is superior to hemoglobin A1c measurement in predicting long-term hypoglycemia in persons with type 2 diabetes. The risk of hypoglycemia associated with treatment intensification may be offset by strategies that reduce glucose variability.

Arch Intern Med. 2004;164:1445-1450
The purpose of this study was to determine if the glucose data obtained by these subjects during the intensive monitoring protocol could be used to identify those at risk for hypoglycemia in the following year when subjects were measuring their blood glucose levels ad libitum. Using SMBG for this purpose requires a theoretical construct in which multiple readings are reduced to parameters that express this risk. Our approach was based on the premise that an individual’s risk of hypoglycemia is a function of his or her mean blood glucose level and its standard deviation. The mean blood glucose reflects the intensity of glycemic control. The standard deviation summarizes the effects of a variety of physiologic and behavioral factors that contribute to glucose variability. The likelihood that any reading will fall below a critical value is increased if the mean glucose is low or its standard deviation is high. This strategy decomposes the risk of hypoglycemia into 2 components that might be targeted by different strategies to minimize its frequency. For example, if this concept is valid, then the increased risk of hypoglycemia caused by intensive glycemic control might be offset by strategies that reduce glucose variability. This concept can be validated by showing that a multivariate model comprised of mean blood glucose and its standard deviation can identify patients with different rates of hypoglycemia in the future. We used a split-sample method in which the model was developed in one randomly selected cohort and then validated in an independent sample followed for up to 1 year.

### METHODS

#### PATIENTS

Potential subjects were identified from computer pharmacy records at the New Mexico VA Health Care System, the Carl T. Hayden VA Medical Center, and the Southern Arizona VA Health Care System. Candidates were randomly sampled using microcomputer-generated random numbers. Patients were eligible for this study if they developed diabetes after the age of 35 years, had no history of diabetic ketoacidosis, took at least 1 injection of a long-acting insulin preparation daily, and did not self-titrate their insulin doses. They were excluded if they had a history of alcoholism or other substance abuse; had chronic liver disease, pancreatic insufficiency, a chronic infectious disease, or an endocrinopathy associated with abnormal glucose homeostasis; had a most recent creatinine value of greater than 3.0 mg/dL (264 μmol/L); received glucocorticoids or immunosuppressive drugs; were being treated with an insulin pump; or had a life-expectancy of less than 1 year.

Eligible subjects were required to have stable diabetes for at least 2 months. Diabetes was considered stable if (1) the daily insulin dose was increased by no more than 10 U or 15% of the baseline dose, whichever was smaller; (2) no new oral medications were prescribed; and (3) the dose of established oral medications was not increased. At entry, subjects completed questionnaires on their disease complications, disabilities, physical activities, treatment regimens, and dietary habits. We also measured height, weight, and HbA1c. A Tosoh (Grove City, Ohio) high-performance liquid chromatography method was used at each study site to measure HbA1c.

Research coordinators trained subjects to correctly use a glucose test meter and observed them taking readings until they could generate reproducible results. The first 218 subjects enrolled in this study were instructed to measure blood glucose levels before breakfast, before lunch, before dinner, and at bedtime for 8 weeks. Subjects returned to the coordinator at 4 and 8 weeks to have their glucose test meters downloaded and provide blood samples for HbA1c. GLUMEAN refers to the mean blood glucose for the 8-week period, while GLUSD refers to the 8-week glucose standard deviation.

After completing this protocol, subjects resumed testing their blood glucose levels in their customary manner. Blood glucose readings were obtained with Comfort Curve test strips and an electronic blood glucose test meter (AccuCheck Complete, Roche Diagnostics, Indianapolis, Ind) capable of storing 1000 readings and their dates and times. Subjects made follow-up visits to the research coordinators at 26, 39, and 52 weeks. At these visits, research coordinators observed subjects performing SMBG, identified all hypoglycemic events captured by the glucose test meters, and measured HbA1c. Hypoglycemia was defined as a blood glucose level of 60 mg/dL or less (<3.33 mmol/L) during follow-up, research personnel made no recommendations or referrals regarding diabetes management and deferred all treatment decisions to the primary care provider. However, investigators notified the primary care provider immediately for all emergency findings.

Informed consent was obtained from all subjects. This protocol, which covered both the initial evaluation and follow-up period, was approved by the institutional review boards at the participating medical centers.

#### STATISTICAL ANALYSIS

The principal hypothesis for this study was that the risk of hypoglycemia could be expressed as a function of mean blood glucose and its variability. The SMBG data were used to develop and validate a multivariate model for estimating their contributions to this risk. Computer-generated random numbers were used to assign approximately 40% of subjects to a derivation set and 60% to a validation set.

**Model Derivation**

Each subject in the derivation set was assigned to 1 of 3 groups based on the number of hypoglycemic events observed during the 8 weeks of intensified SMBG. The first group (group 0) comprised those who did not develop hypoglycemia. The remaining patients were divided into 2 groups depending on whether the number of events was less than or equal to (group 1) or greater than (group 2) the median number of hypoglycemic events in patients who developed hypoglycemia. This classification was necessary because the skewed distribution of hypoglycemic events made it unsuitable for use as a dependent variable in multiple linear regression. The groupings represented a ranking of subjects by hypoglycemia frequency. The group number was therefore treated as an ordinal variable.

Polychotomous logistic regression for an ordinal dependent variable was used to examine the effect of the glucose parameters on group membership. The dependent variable was the group number. The candidate predictor variables were GLUMEAN and GLUSD. The output of this analysis was the probability of membership in all groups exceeding each level of the grouping variable. Because there were 3 such levels (0,
1, and 2), the procedure generated 2 probabilities—the probability of belonging to a group >0 [P(group >0)] and the probability of belonging to a group >1 [P(group >1)]. Note that P(group >0) was the probability of developing at least one episode of hypoglycemia during the 8 weeks of intensified monitoring, whereas P(group >1) was the probability of developing hypoglycemia at a rate exceeding the median frequency.

GLUMEAN and GLUSD were incorporated into the model by a forward-stepping procedure with an α ≤ .05 to enter. The degree of improvement with each step was analyzed by the improvement χ² test, a method that examines the statistical significance of the change in log likelihood. In polychotomous logistic regression, the probability of group membership greater than any level is always a function of the same linear combination of GLUMEAN and GLUSD. P(group >0) was given by the expression

\[ e^{u0} = \text{constant} + k_{\text{GLUMEAN}} \times \text{GLUMEAN} + k_{\text{GLUSD}} \times \text{GLUSD} \]

P(group >1) was given by the expression

\[ e^{u1} = \text{constant} + k_{\text{GLUMEAN}} \times \text{GLUMEAN} + k_{\text{GLUSD}} \times \text{GLUSD} \]

Note that the terms u₀ and u₁ were identical except for their constants. Polychotomous logistic regression generated estimates for the constants, k_{\text{GLUMEAN}} and k_{\text{GLUSD}}. The independent variables were considered statistically significant if the 95% confidence intervals (CIs) for their odds ratios did not include 1.

**Model Validation**

Once the models were developed, it was possible to assign a predicted probability of group membership for each patient in the validation set. These predicted probabilities were calculated by substituting the subject’s GLUMEAN and GLUSD into the expressions for P(group >0) and P(group >1) developed from the derivation sample. We hypothesized that the predicted probability of developing hypoglycemia during 8 weeks of intensified SMBG [P(group >0)] reflected the patient’s risk of hypoglycemia on long-term follow-up (“long-term hypoglycemia”). To test this hypothesis, we used P(group >0) to stratify validation subjects into 4 quartiles of risk. Differences in the occurrence of long-term hypoglycemia across risk strata were tested by χ² analysis. We also hypothesized that the predicted probability of developing hypoglycemia at a rate exceeding the median value during intensified SMBG [P(group >1)] was predictive of long-term hypoglycemia at a rate exceeding the median value (“long-term, high-frequency hypoglycemia”). To test this hypothesis, we used P(group >1) to stratify subjects into quartiles and tested differences in the occurrence of long-term, high-frequency hypoglycemia by χ² analysis.

We compared P(group >0) and HbA1c for their ability to identify patients at risk for long-term hypoglycemia. Because they were continuous variables, the most appropriate method was a receiver operating characteristic (ROC) curve analysis. This analysis calculated the proportions of true positives and false positives as the cut point for a “positive” test was varied across the range of values for the predictor variable. The ROC curve was generated by plotting true-positive vs false-positive rates. The area under the ROC curve measured whether the predictor distinguished subjects who developed long-term hypoglycemia from those who did not. A perfect predictor will have an area under the ROC curve of 1.0, while a predictor with no discriminating ability will have an area of 0.5. The standard error (SE) for the ROC curve area was calculated by the method of Hanley and McNeil. The difference in the ROC curve areas for P(group >0) and HbA1c was considered statistically significant if their 95% CIs did not overlap. The 95% CI for the ROC curve area was considered the mean value ± 1.96 SE. The analysis was repeated to determine if P(group >1) and HbA1c were able to identify patients at risk for long-term, high-frequency hypoglycemia.

All results are reported as mean ± SD. P<.05 was considered statistically significant.

**RESULTS**

We enrolled 218 subjects in the study. Eight withdrew from the intensified monitoring protocol after initially consenting, and 15 were lost to follow-up. Seventy-two (36.9%) of the remaining 195 subjects were randomly assigned to the derivation set, while 123 (63.1%) were assigned to the validation set.

**MODEL DERIVATION**

The mean age of the derivation subjects was 65.0±9.8 years, 11.1% were women, and 26.8% were members of a racial or ethnic minority. The mean body mass index (calculated as weight in kilograms divided by the square of height in meters) was 32.0±5.7, the amount of exercise averaged 55.7±59.0 metabolic equivalent task (MET)-hours per week, and the entry HbA1c was 7.79±1.35%. Thirty-two percent of patients were taking oral hypoglycemic agents, and the mean dose of insulin was 63.5±39.2 U/d. Subjects were asked how many times they had had a low blood glucose value associated with sweating, weakness, anxiety, trembling, hunger, or headache in the month before study entry. Responses were grouped into 5 categories: 31 reported no episodes, 25 reported 1 to 3 episodes, 8 reported 4 to 6 episodes, 2 reported 7 to 12 episodes, and 2 reported more than 12 episodes. Four subjects did not answer this question.

During the 8 weeks of intensified monitoring, the mean blood glucose value was 181±41 mg/dL (10.0±2.8 mmol/L), and the glucose standard deviation was 65±20 mg/dL (3.61±1.17 mmol/L). Sixty-one subjects missed no more than 3 consecutive days of readings. In this group, there was little variation in the weekly glucose mean (range, 175 mg/dL [9.71 mmol/L] to 180 mg/dL [9.99 mmol/L]) or in the weekly glucose standard deviation (range, 57 mg/dL [3.16 mmol/L] to 62 mg/dL [3.44 mmol/L]). For this reason, the 8-week glucose mean (GLUMEAN) and standard deviation (GLUSD) were used for model derivation. Hypoglycemia was detected in 36 subjects (50.0%). The median number of events among the hypoglycemic subjects was 2.5 (interquartile range, 2-5). Kruskal-Wallis one-way analysis of variance by ranks showed that the number of events detected during intensified monitoring was only marginally different (P =.07) across the groups that had reported different hypoglycemic rates at entry. Subjects who became hypoglycemic had a significantly lower 8-week mean blood glucose level (166±31 mg/dL [9.21±1.72 mmol/L] vs 197±45 mg/dL [10.9±2.50 mmol/L]; P =.001). However, no differences were found in glucose standard deviation (64±21 mg/dL [3.55±1.17 mmol/L] vs 67±21 mg/dL [3.72±1.17 mmol/L], respectively; P =.61). GLUMEAN was negatively correlated with the number of blood glucose readings taken during intensified monitoring (r =−.388; P =.002), but no relationship was found for GLUSD. The 36 subjects who did not develop hypoglycemia during intensified SMBG were assigned to group 0, 18 who had 2 or fewer episodes were assigned to group...
There were highly significant differences in the occurrence of long-term hypoglycemia across quartiles defined by $P(\text{group}>0)$. These proportions were 19.4%, 36.7%, 61.3%, and 77.4%, respectively ($P<.001$). The ROC curve analysis showed that $P(\text{group}>0)$ successfully distinguished subjects who eventually developed long-term hypoglycemia from those who did not (area = 0.746 ± 0.046). However, the ROC curve area for $HbA_1c$ (0.549 ± 0.052) was not significantly different from 0.5 and therefore had no prognostic value.

The logistic model also identified subjects who would eventually develop long-term, high-frequency hypoglycemia. The proportions of subjects with this complication in the quartiles defined by $P(\text{group}>1)$ were 0%, 10.0%, 32.3%, and 48.4% ($P<.001$). The ROC area for $P(\text{group}>1)$ was 0.813 ± 0.039, a value significantly greater than the ROC area for $HbA_1c$ (0.540 ± 0.061). Again, the $HbA_1c$ test failed to identify subjects who would develop hypoglycemia exceeding the median frequency.

We also evaluated whether the model could still accurately predict hypoglycemic risk using fewer glucose readings. This analysis was important because 8 weeks of intensified monitoring is impractical for most patients. A total of 119 subjects in the validation set obtained glucose readings for 4 or more days in the first week of intensified monitoring. The mean number obtained was 23.4 ± 5.2. For each of these subjects, GLUMEAN and GLUSD were derived from these data and used in the expressions shown in the Table to recalculate individual probabilities for $P(\text{group}>0)$ and $P(\text{group}>1)$. The cohort was then stratified by each calculated probability into 4 quartiles of risk. On long-term follow-up (mean, 39.8 ± 6.8 weeks), there were highly significant differences in the occurrence of long-term hypoglycemia from the lowest to highest quartile defined by $P(\text{group}>0)$ (33.3%, 30.0%, 55.2%, and 76.7%, respectively; $P = .001$). Likewise, there were significant differences in the occurrence of long-term, high-frequency hypoglycemia across quartiles defined by $P(\text{group}>1)$ (12.9%, 13.8%, 24.1%, and 40.0%, respectively; $P = .04$).

### MODEL VALIDATION

The mean age of the 123 subjects in the validation set was 64.8 ± 10.0 years, 1.6% were women, and 32.8% were members of a racial or ethnic minority group. The mean body mass index was 31.7 ± 5.9, the amount of exercise averaged 65.6 ± 61.5 MET-hours per week, and the entry $HbA_1c$ was 8.24% ± 1.81%. Forty-one percent were taking oral hypoglycemic agents, and the mean dosage of insulin was 64.4 ± 45.4 U/d. During 8 weeks of intensified monitoring, their mean blood glucose was 176 ± 41 mg/dL (9.77 ± 2.28 mmol/L), and their glucose standard deviation was 62 ± 19 mg/dL (3.44 ± 1.05 mmol/L). For each subject, the expressions shown in the Table were used to calculate a predicted probability of hypoglycemia [$P(\text{group}>0)$] and a predicted probability of hypoglycemia at a rate exceeding the median value [$P(\text{group}>1)$]. $P(\text{group}>0)$ was used to stratify the validation set into quartiles; the cut points for this classification were 0.310, 0.545, and 0.769. The process was repeated using $P(\text{group}>1)$. These cut points were 0.084, 0.197, and 0.406.

After completing the intensified monitoring protocol, subjects were followed up for a mean of 39.7 ± 7.1 weeks. During this time, 60 (48.8%) developed hypoglycemia. The median rate of long-term hypoglycemia for these subjects was 0.570 events per 4 weeks (interquartile range, 0.190-1.14). The 28 subjects with hypoglycemia rates exceeding 0.570 events per 4 weeks were considered to have “long-term, high-frequency hypoglycemia.”
level. In addition, this risk can be ascertained with a single week of intensified blood glucose monitoring as long as readings are obtained before meals and at bedtime. Finally, the validation patients were treated by their health care providers in the customary fashion. On long-term follow-up, 24.4% had their oral hypoglycemic agents changed and 97.6% had their insulin preparations switched or insulin doses adjusted. The fact that the model was able to predict hypoglycemic events in subjects on constantly varying treatment regimens suggests that the method is quite robust and minimally affected by treatment changes that occur in routine practice.

Modeling of hypoglycemia risk based on glucose kinetics has certain advantages because the functions provide a method for estimating the risk for different combinations of the glucose variables. For example, our model shows that the estimated risk for at least 1 episode of hypoglycemia in 8 weeks of intensified monitoring is 6% for a subject with a mean glucose of 200 mg/dL (11.1 mmol/L) and a standard deviation of 50 mg/dL (2.78 mmol/L). This risk increases to 60% for a subject with a mean glucose value of 150 mg/dL (8.33 mmol/L) and the same standard deviation but returns to 6% in a patient with a mean glucose value of 150 mg/dL (8.33 mmol/L) and a standard deviation of 20.5 mg/dL (1.14 mmol/L). This analysis suggests that minimizing glucose variability is a plausible method for offsetting the increased risk of hypoglycemia associated with tight glycemic control. However, the method describes changes in hypoglycemia risk from patient to patient as a function of their glucose parameters. It should not be used to forecast a change in the risk within an individual when his or her GLUMEAN or GLUSD is modified.

It seems obvious that patients who are tightly controlled would have a higher risk of hypoglycemia in the future. However, in type 1 diabetes, it has been difficult to demonstrate that HbA1c is an indicator of this risk.6,8 The present study suggests that self-measured blood glucose is a much better predictor of future hypoglycemic events than HbA1c. In addition, a large variation in blood glucose was an independent risk factor that can be ascertained only by frequent blood glucose monitoring. Our study shows that, although HbA1c is the standard measure for glycemic control, it cannot substitute for intensified SMBG in identifying patients who will develop hypoglycemia in the future.

The hypoglycemia risk associated with large fluctuations in blood glucose levels has important clinical implications. This finding suggests that clinicians should attempt to reduce glucose variations attributable to dietary irregularities, mismatching of caloric intake and insulin therapy, medication noncompliance, or exercise during treatment intensification. This concept was exemplified by a randomized clinical trial comparing implantable insulin pumps to multiple daily insulin injections in subjects with type 2 diabetes.12 A reduction in glucose variability achieved by the former method resulted in a significant reduction in the incidence of mild, clinical hypoglycemia. Future studies should be done to validate the concept that reducing glucose variability mitigates the higher rates of hypoglycemia associated with treatment intensification. It should also be noted that our study established GLUMEAN as a predictor, but not necessarily a cause, of hypoglycemia. This distinction is important because the relationship between these 2 variables might have been confounded by certain patient behaviors. For example, a low glucose mean and a high rate of hypoglycemia could be separate manifestations of excessive dieting or exercise. This possibility is supported by our finding that the lowest mean glucose values were found in the patients who were most compliant with our intensive monitoring protocol. Accordingly, we believe that a low mean glucose value in a patient with hypoglycemia should prompt a search for high-risk behaviors. It remains to be proven that allowing the mean glucose level to drift upward will decrease the hypoglycemia rate.

Several studies have shown that parameters derived from glucose monitoring are useful for evaluating the risk of hypoglycemia in type 1 diabetes.6,8,9 The LBGI is derived from low blood glucose readings obtained during routine monitoring. The risk attributable to a single low reading is quantified by logarithmically transforming and quadratically weighting the raw value. The LBGI is the sum of the risks for the low readings divided by the total number of readings. Cox et al6 showed that the LBGI and blood glucose standard deviation, but not HbA1c, were independent predictors of severe hypoglycemia on multivariate analysis. In a larger cohort, this group identified thresholds for LBGI for which there were stepped increases in the odds ratio for future events.8 Only the LBGI and a history of severe hypoglycemia were identified as independent risk factors. Finally, a prospective study9 showed that the LBGI rose, mean blood glucose declined, and blood glucose variance increased in the 24 hours before a hypoglycemic episode. These parameters normalized within 48 hours.

There are significant differences between the LBGI and the probability functions reported in this study. The LBGI was validated against the number of hypoglycemic episodes, while our method was used to stratify patients into risk groups. The end point in the LBGI studies was severe hypoglycemia, while our study used hypoglycemia of any severity. The expressions in this study used all glucose readings obtained from the monitoring protocol and incorporated glucose variability. The LBGI is derived from readings below the midpoint of an acceptable range for blood glucose. Finally, our functions are not based on transformed and weighted data. Nevertheless, both studies show that blood glucose readings from a standardized monitoring protocol are extremely useful for determining the risk of hypoglycemia on long-term follow-up.

Because of our population, our conclusions cannot be extrapolated to women, younger patients, nonveterans, or those taking only oral hypoglycemic agents. In addition, the mechanisms of hypoglycemia may differ for patients who have unstable diabetes, are undergoing treatment intensification, or have comorbidities affecting glucose homeostasis such as chronic liver or renal disease.

In summary, intensified SMBG prospectively identified patients at risk for hypoglycemia and those with an excessive hypoglycemia rate on long-term follow-up. This risk was increased by a low mean blood glucose level or large.
glucose standard deviation and could be expressed as a logistic function involving the 2 parameters. Because the risk associated with lowering mean blood glucose level cannot be avoided during aggressive insulin treatment, clinicians should attempt to reduce glucose variability attributable to behavioral and physiologic factors. Future studies should be done to confirm that periodic intensified SMBG can reduce the risk of hypoglycemia in insulin-treated veterans with type 2 diabetes.

Accepted for publication October 31, 2003.

This work was supported by grant VCR 99-007 from the Health Services Research & Development Service and Veterans Integrated Service Network 18, Department of Veterans Affairs.

Dr Murata is trained and widely published in multivariate techniques and performed the statistical analysis for this article.

The views expressed in this article are those of the authors and not necessarily of the Department of Veterans Affairs.

Correspondence: Glen H. Murata, MD, New Mexico VA Health Care System (111GIM), 1501 San Pedro Dr SE, Albuquerque, NM 87108 (glen.murata@med.va.gov).

REFERENCES


