Symmetrization of the Blood Glucose Measurement Scale and Its Applications

BORIS P. KOVATCHEV, PHD
DANIEL J. COX, PHD
LINDA A. GONDER-FREDERICK, PHD
WILLIAM CLARKE, MD

OBJECTIVE — To introduce a data transformation that enhances the power of blood glucose data analyses.

RESEARCH DESIGN AND METHODS — In the standard blood glucose scale, hypoglycemia (blood glucose, <3.9 mmol/l) and hyperglycemia (blood glucose, >10 mmol/l) have very different ranges, and euglycemia is not central in the entire blood glucose range (1.1–33.3 mmol/l). Consequently, the scale is not symmetric and its clinical center (blood glucose, 6–7 mmol/l) is distant from its numerical center (blood glucose, 17 mmol/l). As a result, when blood glucose readings are analyzed, the assumptions of many parametric statistics are routinely violated. We propose a logarithmic data transformation that matches the clinical and numerical center of the blood glucose scale, thus making the transformed data symmetric.

RESULTS — The transformation normalized 203 out of 205 data samples containing 13,584 blood glucose readings of 127 type 1 diabetic individuals. An example illustrates that the mean and standard deviation based on transformed, rather than on raw, data better described subject's blood glucose distribution. Based on transformed data: *1*) the low blood glucose index predicted the occurrence of severe hypoglycemia, while the raw blood glucose data (and glycosylated hemoglobin levels) did not; 2) the high blood glucose index correlated with the subjects' glycosylated hemoglobin ($r = 0.63$, $P < 0.001$); and 3) the low plus high blood glucose index was more sensitive than the raw data to a treatment (blood glucose awareness training) designed to reduce the range of blood glucose fluctuations.

CONCLUSIONS — Using symmetrized, instead of raw, blood glucose data strengthens the existing data analysis procedures and allows for the development of new statistical techniques. It is proposed that raw blood glucose data should be routinely transformed to a symmetric distribution before using parametric statistics.

Blood glucose regulation in individuals with type 1 diabetes is influenced by many factors, resulting in substantial blood glucose fluctuations. The measurement of these fluctuations is often the object of statistical description and data analysis. Most statistical techniques, however, rely on assumptions about the shape of the underlying distribution of the data. Such analyses confront a problem that originates in the blood glucose measurement scale: the hyperglycemic range (blood glucose, >10 mmol/l) is much wider than the hypoglycemic range (blood glucose, <3.9 mmol/l), and the target blood glucose range (blood glucose, 3.9–10.0 mmol/l) defined by the Diabetes Control and Complications Trial (1) is not centered in the whole possible blood glucose range (assuming 1.1–33.3 mmol/l). In other words, there is no blood glucose value that would be both a "clinical" and numerical center of the scale. A clinical center should be a safe euglycemic value, while the numerical center of the standard blood glucose scale is ~17 mmol/l, well inside the hyperglycemic range.

Three examples will clarify this point. First, suppose that based on multiple blood glucose readings, the mean blood glucose concentration and standard deviation are reported for a subject. This describes the data accurately only if the blood glucose readings are symmetrically distributed around the mean. If the distribution is skewed, the description is not accurate. Second, suppose that the means of a series of blood glucose observations for subjects A and B are compared to assess the subjects' glycemic control. However, subjects with similar mean blood glucose concentrations may have similar glycosylated hemoglobin levels, but they may not necessarily have a similar risk for severe hypoglycemia. There is a simple computational reason for this: The average blood glucose concentration is not sensitive to fluctuations of low or high blood glucose, since these fluctuations balance each other out. Even if we consider the mean together with the standard deviation of each series, the problem is not solved, since it requires the symmetry of the blood glucose measurements. Third, suppose that a treatment designed to keep the blood glucose levels of a subject with type 1 diabetes in a target range is tested for effectiveness. The data are multiple self-monitoring of blood glucose (SMBG) readings. To quantify the overall treatment effect, it is necessary to compute a measure that is equally sensitive to low and high blood glucose concentrations. In addition, if it is suspected that the subject's improvement differs in the low versus high blood glucose range, it would be a problem to make these two ranges statistically comparable.

In this paper, we propose a logarithmic-type transformation that: *1*) makes the transformed blood glucose scale symmetric around zero, thus defining a clinical and numerical center point, which makes 6.25 mmol/l the center of the blood glucose scale and the transformed blood glucose readings distribution normal; and 2) serves as a basis for defining the blood glucose risk indexes (2,3), which, given multiple SMBG readings, predict the subjects' glycosylated hemoglobin levels and the likelihood for severe hypoglycemia.

RESEARCH DESIGN AND METHODS

Subjects

The inclusion criteria for the 127 partici-

From the University of Virginia Health Sciences Center, Center for Behavioral Medicine Research, Charlottesville, Virginia.

Address correspondence and reprint requests to Boris P. Kovatchev, PhD, University of Virginia Health Sciences Center, Center for Behavioral Medicine Research, Blue Ridge Hospital, Drawer F, Charlottesville, VA 22901. E-mail: bpk2u@virginia.edu.

Received for publication 7 March 1997 and accepted in revised form 15 July 1997.

Abbreviations: BGAT, blood glucose awareness training; SMBG, self-monitoring of blood glucose.

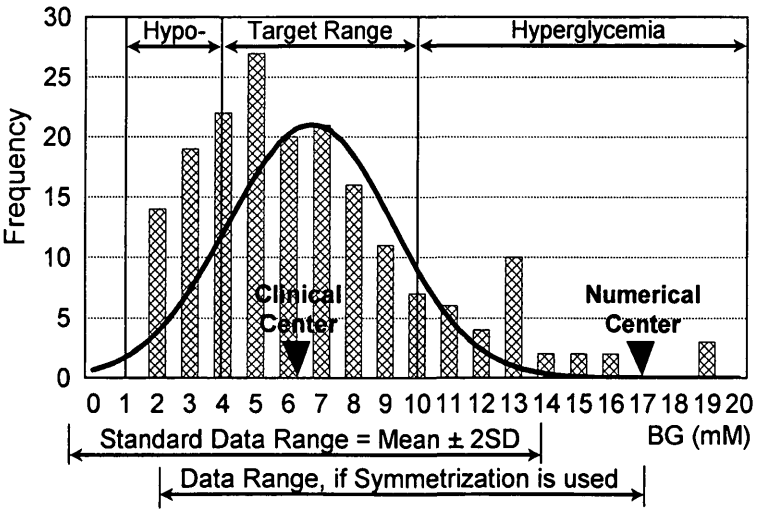

Figure 1—*Typical blood glucose data distribution for a subject with type 1 diabetes.*

pants were a duration of diabetes of at least 2 years, insulin therapy since the time of diagnosis, and the routine performance of SMBG with a blood glucose meter more than twice a day. There were 50 male and 77 female subjects enrolled in the study. The mean age was 38.1 ± 10.0 years; the mean duration of diabetes, 19.6 ± 10.3 years; the mean units of insulin per day, 38.5 ± 15.7 U; and the mean glycosylated hemoglobin, 10.05 ± 1.9%. The normal range for the glycosylated hemoglobin assay was 4.4–6.9%.

Procedure

The subjects were recruited through newsletters, notices posted in diabetes clinics, and direct physician referral. All subjects attended orientation meetings, were informed about the study, signed consent forms, and were instructed to measure their blood glucose concentrations three to five times a day for 1 month. An initial group of 78 subjects performed this procedure twice, separated by a 6-month waiting period. The other 49 subjects performed the procedure once. This resulted in 205 data samples with a total of 13,584 measurements (66 per sample). Lifescan One Touch II memory meters were used by all subjects, and the data were electronically transferred for analysis.

Blood glucose scale symmetrization

We assume a possible blood glucose range of 1.1–33.3 mmol/l and a target blood glucose range of 3.9–10.9 mmol/l (1). To derive a data transformation formula that makes the standard blood glucose measurement scale symmetric, we must "expand" the hypoglycemic range, "squeeze" the hyperglycemic range, and place the target blood glucose range in a central location around zero. This notion leads to the following assumptions: the transformed possible blood glucose range to be symmetric around zero (Assumption 1), and the transformed target blood glucose range to be symmetric around zero (Assumption 2). Having Assumptions 1 and 2, we derived a transformation that symmetrizes the blood glucose scale:

$$\text{transformed blood glucose} = 1.794 \cdot ([\log\{BG\}]^{1.026} - 1.861),$$

where BG is the blood glucose concentration measured in millimoles per liter. If a scale in milligrams per deciliter is used for the blood glucose measurements, the equation is:

$$\text{transformed blood glucose} = 1.509 \cdot ([\log\{BG\}]^{1.084} - 5.381)$$

The mathematical description of the equations and their solutions is given in the APPENDIX. Note that no data are used for parameter evaluation. Based on this transformation, we developed the low and high blood glucose risk index of a subject (2,3). The risk values range from "zero risk" at a blood glucose concentration of 6.25 mmol/l to a "maximum risk" of 100 at 1.1 and 33.3 mmol/l (see APPENDIX).

RESULTS

Blood glucose scale transformation

To verify our data transformation, it was applied to all 205 SMBG data sets. For 203 data samples, the hypothesis that the transformed data fit a normal distribution was confirmed with a maximal two-tailed probability of 0.97. Only two hypotheses were rejected. The Kolmogorov-Smirnov test was used with a P value of 0.005. (Note that with more than 200 tests, this P value ensures a very conservative hypothesis acceptance range.)

Example 1. Figure 1 presents a typical blood glucose data distribution for a subject with type 1 diabetes (186 readings from his memory meter). The distribution is skewed and the superimposed normal density poorly describes the data. The latter is confirmed by a Kolmogorov-Smirnov test ($P < 0.05$). The average blood glucose

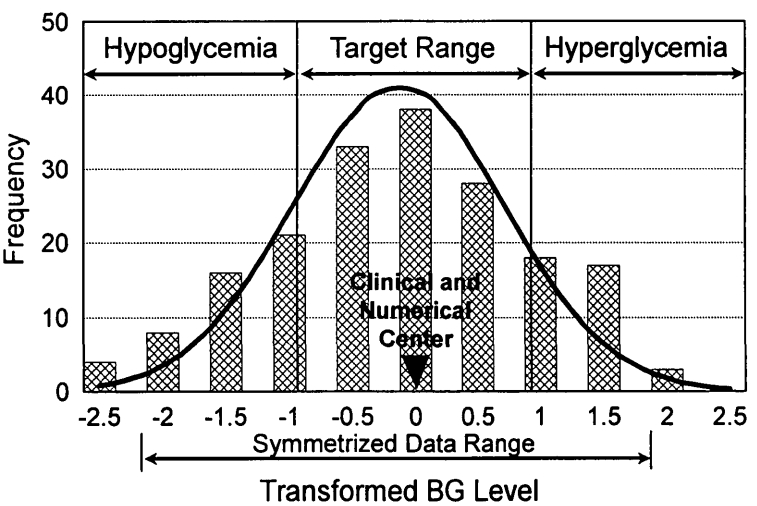

Figure 2—*Histogram of Fig. 1 data following application of data transformation.*

Table 1—*Blood glucose (BG) data for two subjects with type 1 diabetes*

	Average BG (mmol/l)	Glycohemoglobin (%)	Low BG index	Episodes of severe hypoglycemia at 12 months (*n*)
Subject 1	8.8*	8.9	0.4*	0
Subject 2	12.3*	14.9	3.1*	12

*$P < 0.001$. BG, blood glucose.

concentration for this subject is 6.7 ± 3.6 mmol/l, and his standard data range (mean ± 2 SD) is from −0.5 to 13.9. Theoretically, with perfect data, 2.3% (or four readings for this sample size) should fall below the lower bound, and another four readings should fall above the upper bound of that range. However, no readings could possibly fall below its lower bound, simply because it is negative, and nine readings (or 4.8%) fall above the upper bound (i.e., the standard range is substantially shifted left). This discrepancy is due to the asymmetry of the blood glucose data.

Figure 2 is a histogram of the Fig. 1 data after applying our data transformation. The Kolmogorov-Smirnov test results in an almost perfect normal distribution ($P = 0.92$). The mean of the transformed data is −0.13 ± 1.02. By reversing the transformation, we find that the data range in the original blood glucose scale is from 1.9 to 17 mmol/l (Fig. 1). Four readings fall below the lower bound, and another three readings fall above the upper bound of that interval.

Blood glucose indexes

The low and high blood glucose indexes, which quantify the number and extent of the low and high blood glucose readings (2,3; APPENDIX), were reliable over 6 months for the 78 subjects who performed two data collection sessions with a 6-month waiting period. The test-retest correlations were 0.68 and 0.60, respectively ($P < 0.001$). The high blood glucose index was positively correlated with subjects' glycosylated hemoglobin values ($r = 0.63$, $P < 0.001$).

Example 2. The low blood glucose index was associated with severe hypoglycemia, even when the average blood glucose and glycosylated hemoglobin indicated otherwise. Table 1 presents data for two subjects with type 1 diabetes. The second patient had significantly higher average blood glucose concentrations ($P < 0.001$), which went along with his higher glycosylated hemoglobin values. Given the average blood glucose and glycohemoglobin readings, we can conclude that the first subject was at higher risk for severe hypoglycemia, since severe hypoglycemia is associated with lower glycosylated hemoglobin levels (1). However, the low blood glucose index was significantly higher for the second subject ($P < 0.001$), leading to the opposite conclusion, that the second subject was at much higher risk for severe hypoglycemia. Indeed, the first subject reported no severe hypoglycemic episodes in the past year, while the second subject reported 12 episodes.

Example 3. Re-analysis of previously reported data (3) indicated that blood glucose awareness training (BGAT), which enhances the ability of subjects with type 1 diabetes to recognize their blood glucose level, also influences the distribution of their blood glucose measurements. Table 2 summarizes the results. The posttreatment blood glucose index (low plus high) was lower ($P < 0.05$), indicating that BGAT did significantly reduce the extreme blood glucose readings without changing glycosylated hemoglobin levels.

CONCLUSIONS — Our transformation reshapes the blood glucose scale, introducing a numerical central point that is also clinically meaningful. This could have the following clinical and research applications.

First, most parametric statistical tests have normality (or at least symmetry) assumptions that are usually violated when blood glucose data are analyzed. Our example shows that even the computation of a simple data range can be incorrect, if the symmetry of the data is violated. An analysis that is carried out with normalized blood glucose data will ensure greater validity and sensitivity of the results.

Second, clinicians and clinical researchers assume that blood glucose variability is a predictor of severe hypoglycemia. Now we can quantify this notion. The idea behind the blood glucose risk indexes is to capture in a single number the frequency and extent of the blood glucose fluctuations. Thus, the low blood glucose index will be higher for a subject with higher percentage of low blood glucose readings or for a subject who has more extreme hypoglycemic episodes. The high blood glucose index works in the same way at the high end of the blood glucose scale, and this is consistent with its positive correlation with glycosylated hemoglobin values.

Third, the transformed data can give a correct assessment of the risk for severe hypoglycemia, even when the standard measures indicate exactly the opposite. In our previously published data (2), glycosylated hemoglobin values were not predictive of future severe hypoglycemia episodes, although a relationship with the rate of severe hypoglycemia per 100 patient-years has been reported by others (1). In contrast, the low blood glucose index accounted for 46.5% of the variance of the severe hypoglycemia episodes in the subsequent 6-month period (2).

Fourth, re-analysis of previous data (3) indicates that our risk indexes are capable of capturing group and/or treatment effects that would be missed by the standard glycemic control measures.

Finally, the parameters of our transformation are estimated on the basis of widely accepted clinical standards (Assumptions 1 and 2), which have been endorsed by the Diabetes Control and Complications Trial (1). Thus, the transformation can be directly applied to any data sample, without parameter reestimation. In contrast, the standard skewness corrections (4) are based on empirical parameter estimation and, hence, are tied to a particular data sample. The formulas are simple enough to be incorporated as a data transformation in any statistical package used for data analysis.

Table 2—*Summary of results of BGAT*

	Average BG (mmol/l)	Glycohemoglobin (%)	BG index (low plus high)
Pre-BGAT	9.3	10.3	13.8*
Post-BGAT	9.3	10.2	13.2*

*$P < 0.05$. BG, blood glucose.

Acknowledgments— This study was supported by National Institutes of Health grants RO1 DK51562 and RO1 DK 28288 and by a grant from Lifescan (Milpitas, CA). The authors express their appreciation for the extremely useful comments made by Professor Robert Abbott of the University of Virginia.

APPENDIX

Deriving the skewness correction transformation from clinical assumptions

Since blood glucose is a concentration (of glucose in the blood), we can expect that an appropriate data-normalizing transformation would be of logarithmic type. A similar conclusion can be made if the classic Box-Cox skewness correction transformation (4) is applied to satisfy Assumptions 1 and 2. Thus, as a starting point, we used a two-parameter logarithmic-type transformation:

$$U(BG; \alpha, \beta) = (\log[BG])^{\alpha} - \beta$$

We added a scaling parameter γ and fixed the minimal and maximal transformed blood glucose values at $\sqrt{10}$, a convenient value for defining our risk function. Assumptions 1 and 2 led to the following equations:

$$(\log[33.3])^{\alpha} - \beta = -([\log\{1.1\}]^{\alpha} - \beta) \quad \text{(Assumption A1)}$$

$$(\log[10.0])^{\alpha} - \beta = -([\log\{3.9\}]^{\alpha} - \beta) \quad \text{(Assumption A2)}$$

$$\gamma \cdot ([\log\{33.3\}]^{\alpha} - \beta) = -\gamma \cdot ([\log\{1.1\}]^{\alpha} - \beta) = 10^{1/2} \quad \text{(Scaling parameter)}$$

These three equations are easily reduced to a single nonlinear equation for the parameter α. When solved numerically under the restriction $\alpha > 0$, it gives α = 1.026, β = 1.861, and γ = 1.794. Given that 33.3 mmol/l = 600 mg/dl, 1.1 mmol/l = 20 mg/dl, 10 mmol/l = 180 mg/dl, and 3.9 mmol/l = 70 mg/dl in the milligrams-per-deciliter scale, the equations above produce α = 1.084, β = 5.381, and γ = 1.509. Thus, the entire blood glucose range is transformed into the symmetric interval (−3.2, 3.2). The target blood glucose range is transformed into the symmetric interval (−0.9, 0.9). Zero is the numerical center of the transformed scale that corresponds to 6.25 mmol/l, a euglycemic value that marks the clinical center of the blood glucose scale.

The blood glucose risk function (3) is defined as $R(BG) = 10 \cdot U(BG)^2$. The low blood glucose risk is defined as LR(BG) = R(BG), if the blood glucose concentration is <6.25 mmol/l, and 0 otherwise. The high blood glucose risk is defined as HR(BG) = R(BG), if the blood glucose concentration is ≥6.25 mmol/l, and 0 otherwise. Given multiple SMBG readings, the low and high blood glucose index are computed as the means of LR(BG) and HR(BG), respectively (2,3).

References

1. The Diabetes Control and Complications Trial Study Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in IDDM. *N Engl J Med* 329:977–986, 1993
2. Cox DJ, Kovatchev BP, Julian DM, Gonder-Frederick LA, Polonsky WH, Schlundt DG, Clarke WL: Frequency of severe hypoglycemia in IDDM can be predicted from self-monitoring blood glucose data. *J Clin Endocrinol Metab* 79:1659–1662, 1994
3. Cox DJ, Gonder-Frederick LA, Polonsky WH, Schlundt DG, Julian DM, Clarke WL: A muticenter evaluation of blood glucose awareness training-II. *Diabetes Care* 18: 523–528, 1995
4. Box GEP, Cox DR: An analysis of transformations (with discussion). *J Royal Stat Soc, Series B* 26:211–252, 1964