What Do Statistics Really Tell Us About the Quality of the Data from Self-monitoring of Blood Glucose?

R. F. Dedrick, W. K. Davis

Michigan Diabetes Research and Training Center, University of Michigan Medical School, Ann Arbor, USA

Recently a number of studies have examined the quality of the data obtained from various systems used in the self-monitoring of blood glucose. Many of these studies have used parametric statistical techniques such as the Pearson product–moment correlation (*r*) and linear regression to evaluate the errors associated with self-monitoring results. These statistical methods, while well known and easily computed on modern computers, are often inappropriate for evaluating either the amount of error associated with self-monitoring or the clinical significance of these errors. For example: 1. a correlation of 1.00 does not necessarily mean that the measurements from a self-monitoring system agree with the true values and are without error; 2. a correlation close to 0.00 does not necessarily mean that the measurements from self-monitoring differ widely from the true values and possess large amounts of error; 3. a slope of 1.0 and a *y*-intercept of 0.0 in a linear regression equation do not necessarily mean that the self-monitoring measurements agree with the true values; 4. a slope and *y*-intercept that deviate significantly from 1.0 and 0.0, respectively, do not necessarily mean that such measurements differ widely from the true values. The present paper illustrates some of the limitations and common misconceptions concerning these statistics, and shows that a reliance on these techniques alone can, in certain circumstances, lead to misleading estimates of the amount of error associated with self-monitoring systems and inappropriate descriptions of the clinical significance of these errors. We would wish to discourage the use of these statistics for evaluating the clinical significance of the errors in self-monitoring results, and encourage the use of more appropriate analyses such as error grid analysis.

KEY WORDS **Self-monitoring Blood glucose Diabetes Measurement error Statistical analysis**

Introduction

In the last 8 years, as self-monitoring of blood glucose has become a more important part of the management of diabetes, there has emerged a growing concern about the quality of the data obtained using this procedure outside the clinical laboratory.[1] This concern has resulted in a number of studies that have examined the quality of the self-monitoring data obtained by a variety of individuals (children, adolescents, and adults with diabetes; parents and health professionals) using various reagent strips and meters.[2–17]

Most of these studies have examined the quality of the self-monitoring data by comparing the blood glucose measurements obtained from the self-monitoring system with the measurements determined using a laboratory method (Beckman Glucose Analyzer II, YSI Model 23A Glucose Analyzer, and others). The laboratory method is used to represent the 'gold standard' or 'true' measure of blood glucose, and the quality of the data is evaluated by examining the extent to which the patient-generated measurements deviate from the 'true' values. The deviations from the 'true' values are viewed as errors in the self-monitoring system and may be due to random factors (chance variation), systematic factors (bias), or a combination of random and systematic factors. Random error increases the imprecision of the measurements, while systematic error (either constant or proportional) increases the inaccuracy of the measurements.[18]

While the design of these studies is straightforward, there has been some controversy over what methods should be used to analyse the data resulting from them. Many studies have used statistical methods such as the Pearson product–moment correlation (*r*) and linear regression to evaluate the error associated with self-monitoring, despite the fact that a number of investigators have noted the limitations of these methods.[19–21] Clarke *et al.*, for example, have argued that these statistics evaluate the deviations of the patient-generated measurements from their true values 'in ways that may not be clinically useful and therefore make it difficult to evaluate the clinical significance of a particular product or method'.[21]

In place of these statistical methods, Cox *et al.*[19] have developed a systematic and comprehensive graphical display technique for evaluating the errors associated

Correspondence to: Dr Robert F. Dedrick, University of Michigan Medical School, 1500 E. Medical Center Drive/G1113 Towsley, Box 0201, Ann Arbor, MI 48109, USA.

0742–3071/89/030267–07 $05.00

with self-monitoring. This technique, which is called error grid analysis, displays a scatterplot of the self-monitoring results against the laboratory measurements and classifies the errors in the patient-generated data in terms of their clinical significance (Figure 1). Clinical significance is determined by considering the percent error (deviation) between the patient-generated and true measurements in relation to the patient's true glucose level. The major principle underlying this display technique is that two patient-generated measurements that deviate by the same amount from their true values (e.g. 50 %) may have different clinical meanings depending on the patients' true glucose levels. For example, a 50 % underestimate of a true value of 3.3 mmol l^{-1} will not make any difference in terms of the treatment actions taken by a patient. Under both circumstances (an incorrect blood glucose reading of 1.7 mmol l^{-1} or a correct reading of 3.3 mmol l^{-1}) the patient will take corrective treatment action to raise the blood glucose concentration. However, a 50 % underestimate when the true value is 16.7 mmol l^{-1} will result in the patient not taking action to lower a clinically elevated blood glucose level.

Error grid analysis defines five major zones or categories of errors that range from Zone A to Zone E (note that Cox and colleagues'[19] definition of clinical significance reflects the practices and philosophy of the University of Virginia Medical Center, but can be adjusted to meet the needs of other institutions). Zone A in error grid analysis represents clinically accurate measurements. These measurements are within 20 % of the laboratory measurements or are in the hypoglycaemic range (< 3.9 mmol l^{-1}) when the true glucose levels are also in the hypoglycaemic range.[20] Zone B represents measurements that deviate from the laboratory values by more than 20 %; these measurements, however, would lead to either no change in treatment or benign treatment changes (treatment changes that would maintain actual blood glucose levels within the 3.9 mmol l^{-1} to 10 mmol l^{-1} range). For example, a patient-generated measurement of 7.8 mmol l^{-1} when the true value is 5.6 mmol l^{-1} is in error by 39%, but this error would not lead to any corrective treatment action. Zone C refers to measurements that deviate from the laboratory by more than 20 % and 'would lead to unnecessary corrective treatment errors'.[20] These unnecessary corrective treatment errors might result in the actual blood glucose levels falling below the 3.9 mmol l^{-1} level or rising above 10 mmol l^{-1}. Zone D involves 'potentially dangerous failures to detect and treat extreme reference blood glucose'.[20] These errors occur when the patient-generated measurements are within the 3.9 to 10 mmol l^{-1} range but the true glucose levels are outside this target range. Zone E represents measurement errors 'that would result in erroneous self-treatment'.[20] Patient-generated measurements in this zone 'are opposite to the reference values, and corresponding treatment decisions would therefore be opposite to that called for'.[21] For example, a patient-generated measurement of 2.2 mmol l^{-1} when the true value was 11.1 mmol l^{-1} would lead the patient to take action to increase a clinically low blood glucose level, when in fact the patient should be taking action to lower a clinically elevated blood glucose level.

The errors in the error grid system represent the total deviation of the patient-generated measurement from the true values, and are not subdivided into random and systematic components. This approach to conceptualizing the errors associated with self-monitoring systems reflects the clinical orientation of the error grid technique and is consistent with the view held by Westgard *et al.*[22] that an examination of the total error of a system is more medically useful for evaluating the performance of the system outside the laboratory. These authors note that 'the physician thinks rather in terms of the total analytic error, which includes both random and systematic components. From his [physician] point of view, all types of analytic error are acceptable as long as the total analytic error is less than a specified amount. After all, it makes little difference to the patient whether a laboratory value is in error because of random or systematic analytic error'.[22]

In using error grid analysis, the error associated with the self-monitoring system is evaluated by calculating the percentage of self-monitoring measurements that fall within each of the five zones. Figure 1 illustrates the use of error grid analysis with the hypothetical data contained in Table 1, and displays the critical boundaries that define the clinical significance of the patient-generated errors. As can be seen in this graph, 33 % of the self-monitoring measurements were in Zone C, 33 % were in Zone D, and 33% were in Zone E. None of the patient-generated values were clinically accurate or acceptable.

While error grid analysis represents a powerful tool for evaluating the clinical significance of the errors associated with self-monitoring and provides a means of clarifying the information provided by statistical methods, it has not been widely used in published research reports examining self-monitoring data generated outside the laboratory. In general, many investigators have continued to rely solely on statistical methods to evaluate the errors associated with self-monitoring. For example, at the recent Consensus Development Conference on Self-Blood Glucose Monitoring (November 1986), six of the 13 abstracts that were presented used either correlation or regression techniques to evaluate the self-monitoring systems, and none of the abstracts used error grid analysis. One reason for the continued reliance on statistical methods to evaluate errors in self-monitoring may be that, while Cox *et al.*[19] and Clarke *et al.*[21] have made an excellent case for using error grid analysis, these individuals and others have not gone far enough in clarifying the limitations and potential statistical problems with the Pearson product–moment correlation and linear regression techniques. Clarke *et al.*, for example, criticized the correlation coefficient as a measure of the clinical accuracy of a self-monitoring system, noting that the '*r* value can be close to unity for large

Table 1. Hypothetical glucose values (mmol l^{-1}) from a self-monitoring device and a laboratory method illustrating a Pearson product–moment correlation of 1.00

Laboratory reference value (mmol l^{-1})	Self-monitoring value (mmol l^{-1})	% Deviation	Error grid analysis zone	Clinical significance[a]
1.11	5.00	350	D	Dangerous failure to treat
1.67	7.50	350	D	Dangerous failure to treat
2.78	12.50	350	E	Erroneous treatment
3.33	15.00	350	E	Erroneous treatment
4.44	20.00	350	C	Unnecessary correction
5.55	25.00	350	C	Unnecessary correction

[a]Clinical significance is based on error grid analysis.[19]

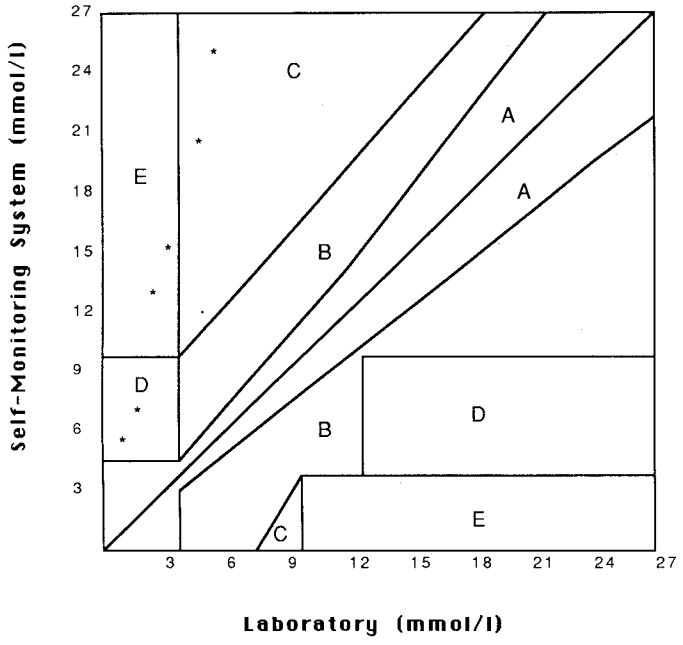

Figure 1. Example of error grid analysis using the data in Table 1. The target range of desirable glucose values established by Cox *et al.* is 3.9 mmol l^{-1} to 10.0 mmol l^{-1}. The lower level Zone C (zone below the diagonal) ranges from 7.2 mmol l^{-1} to 10.0 mmol l^{-1}. Lower level Zone D begins at 13.3 mmol l^{-1}. Lower level Zone E begins at 10.0 mmol l^{-1}

sets of data when individual data points differ by large amounts. In addition, correlation coefficients that evaluate the entire range of blood glucose may misrepresent the true relationship between subsets of data'.[21] These same authors similarly criticized the linear regression technique noting that 'a slope that approaches unity cannot always predict the relationship between two specific data points'.[21] While these statements are correct, they may not go far enough in explaining the nature of correlation and regression analysis and the conditions under which these techniques may provide misleading information about the amount and clinical significance of the errors associated with the self-monitoring system.

Given the potential dangers in relying solely on statistical techniques to evaluate the error associated with self-monitoring, this paper illustrates some of the limitations and common misconceptions concerning these statistics, and demonstrates that under certain conditions these statistics provide misleading estimates of the amount of error present in the self-monitoring system. In addition, this paper illustrates that these techniques provide little information about the clinical significance of the self-monitoring errors. Although it is intended primarily for investigators examining the performance of self-monitoring systems in clinical applications, many of the issues discussed are relevant to analytical evaluations conducted in the clinical laboratory.

Pearson Product–Moment Correlation

A number of studies have incorrectly used the Pearson product–moment correlation (r) to measure the extent to which the measurements from self-monitoring systems agree with the true values of blood glucose.[3,5,6,9,15,16] Many of the authors of these studies have incorrectly assumed that correlations that approach (or are equal to) a positive, perfect correlation ($r = 1.00$) indicate a small amount of error in the self-monitoring system, while correlation coefficients that approach 0.00 indicate a large amount of error.

These assumptions are incorrect because the Pearson r does not provide an estimate of the total amount of error associated with a self-monitoring system, but rather only provides information about the random error (imprecision) influencing the system. Thus, an r of 1.00 does not necessarily mean that the measurements from the self-monitoring system are the same, or even close to the

Table 2. Hypothetical glucose values (mmol l^{-1}) from a self-monitoring device and a laboratory method illustrating a weak correlation ($r = 0.25$)

Laboratory reference value (mmol l^{-1})	Self-monitoring value (mmol l^{-1})	% Deviation	Error grid analysis zone	Clinical significance
5.3	5.5	3.8	A	Clinically accurate
5.4	5.5	1.9	A	Clinically accurate
5.5	5.3	3.6	A	Clinically accurate
5.5	5.8	5.5	A	Clinically accurate
5.6	5.7	1.8	A	Clinically accurate
5.6	5.6	0.0	A	Clinically accurate
5.6	5.7	1.8	A	Clinically accurate
5.7	5.6	1.8	A	Clinically accurate
5.7	5.5	3.5	A	Clinically accurate
5.7	5.6	1.8	A	Clinically accurate

measurements obtained using the laboratory method. For example, the correlation for the hypothetical data presented in Table 1 is 1.00 even though each of the values from the self-monitoring device deviates from the reference by 350 %, and each represents a potentially dangerous clinical error as defined in the error grid system.

This result can occur because the Pearson *r* is a statistic designed to measure the extent to which pairs of measurements are associated, not the extent to which measurements agree. While the terms association and agreement are frequently used synonymously in everyday language, these terms have very different and precise meanings in statistics.[23] Association represents the extent to which pairs of measurements go together according to some mathematical rule or function, while agreement represents the extent to which pairs of measurements are identical. According to these definitions, agreement may be viewed as a special case of association. What this means is that if two measurements agree they will be associated, but two measurements that are associated will not necessarily agree. The implication of this fact is that if measurements from a self-monitoring system differ from a reference (laboratory measurement) by a proportional amount (e.g. a systematic error of 350 %) or a constant amount (e.g. a systematic error of 350 units) the correlations will be 1.00 because these measurements, although not identical, will exhibit perfect association. This association, however, does not mean that there is an absence of error in the self-monitoring system.

The inappropriateness of the Pearson *r* as a measure of the error associated with a self-monitoring system is further illustrated by the hypothetical data in Table 2. The Pearson correlation for these data is 0.25 despite the fact that each of the measurements from the self-monitoring system is almost identical to the measurements obtained in the laboratory (the largest percent error was 5.5 %), and each is 'clinically accurate' as defined by error grid analysis. This result can occur because the Pearson *r* is influenced by the variability of the measurements that are correlated. Measurements that have little variability will not correlate strongly, and thus if the range (variability) of one or more of the measurements is restricted (e.g. examining the blood glucose values in the 5.3–5.7 mmol l^{-1} range) the correlation will decrease. Figure 2 illustrates what happens to the correlation coefficient when the range of glucose values under investigation is increased. The correlation for the range of glucose values from 2.0 mmol l^{-1} to 14.0 mmol l^{-1} is equal to 0.98, while the correlations in the 2.0–4.0 mmol l^{-1}, 6.0–8.0 mmol l^{-1}, and 12.0–14.0 mmol l^{-1} ranges are each equal to 0.30.

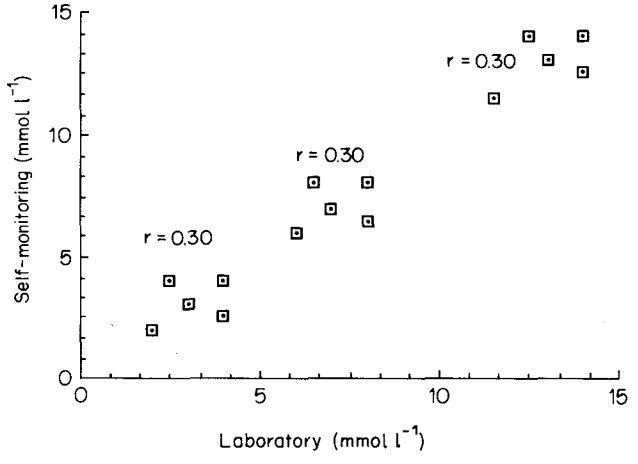

Figure 2. Scatterplot illustrating the influence of variability of blood glucose estimates on the correlation coefficient. Correlation for the full range is 0.98

While it might be argued that investigators examining the performance of self-monitoring systems (both within the laboratory and in clinical practice) should include a representative range of blood glucose values in their performance assessments, currently there is no agreement on what constitutes a 'representative' range of blood glucose values. The implication of this fact is that investigators will use different ranges of values, thus making the interpretation of the Pearson *r* problematic.

These examples illustrate that the Pearson r does not provide information about the amount of systematic error associated with a self-monitoring system. Furthermore, while the Pearson correlation provides some information about random error (imprecision) this statistic is difficult to interpret because the size of the correlation is influenced by the range of the measurements. Given these facts it is clear that this statistic should not be used to quantify the error associated with a self-monitoring system, and consequently should not be used to evaluate the clinical significance of the errors that occur in using self-monitoring.

Linear Regression

Least squares linear regression has also been widely used to evaluate the quality of the data obtained from various self-monitoring systems.[5,6,9,12,14,16] These studies have examined the amount of error associated with self-monitoring by fitting a linear regression equation to the blood glucose values obtained using a given method of self-monitoring and the corresponding blood glucose measurements obtained in the laboratory. The linear regression line, represented by the equation $y = ax + b$, where x is the independent variable ('gold standard' or laboratory measurement), y is the predicted value for the dependent variable (self-monitoring method under evaluation), b is the y-intercept, and a is the slope of the regression line, is determined using the least squares criterion. The least squares criterion involves selecting the slope and y-intercept so that the regression line provides the best fit of the data (i.e. the sum of the squared errors around the regression line is as small as possible).

Once the regression line is fitted to the data, the quality of the data from the self-monitoring system has generally been evaluated by examining the values of the slope and y-intercept. Many investigators have incorrectly assumed that a slope of 1.0 and a y-intercept of 0.0 (i.e. a line that goes through the axis at the origin and forms a 45 degree angle) indicate that the glucose values determined using a given method of self-monitoring are identical to the values determined in the laboratory and are without error. In addition, it has been incorrectly assumed that a slope and/or y-intercept that differ from 1.0 and 0.0, respectively, indicate that there are major discrepancies (error) between the measurements obtained using a self-monitoring device and those obtained using the laboratory method.

The flaw in the first assumption is that while a self-monitoring system that has no error will have values of 1.0 and 0.0 for the slope and y-intercept, respectively, a slope of 1.0 and a y-intercept of 0.0 do not necessarily mean that the measurements from the self-monitoring system are the same, or even close to the measurements obtained using a laboratory method. Table 3 illustrates a case in which the regression line has a slope of 1.0 and a y-intercept of 0.0 despite the fact that the self-monitoring measurements and the laboratory measurements are not identical. This result can occur because the slope and y-intercept do not provide an estimate of the total error associated with self-monitoring, but rather only provide information about the amount of systematic error (inaccuracy) influencing the system. In this example, the discrepancies between the self-monitoring measurements and the true values are due to random factors (imprecision) and thus the values of the slope and y-intercept, which are not sensitive to these errors, do not provide information about the level of agreement between the patient-generated values and the true values. In order to obtain a more complete description of the errors influencing the self-monitoring system it would be necessary to compute the standard error of estimate for the regression line, in addition to determining the slope and y-intercept. The standard error of estimate is the standard deviation of the points around the regression line and provides a quantitative estimate of the amount of random error influencing the self-monitoring system. The value of the standard error of estimate for this example is 3.1 mmol l^{-1} (standard deviation of the self-monitoring values is 4.2 mmol l^{-1}) which would indicate that there is a relatively large amount of random error in the self-monitoring measurements. This value, however, like the slope and y-intercept, provides little information to the clinician about the clinical significance of the errors in the self-monitoring system. In contrast, the results of error grid analysis indicate that 8 out of the 10 errors in the patient-generated data are benign errors (Zone B) while only one measurement represents a dangerous clinical error (i.e. failure to detect and treat error, Zone D).

The assumption that the slope and y-intercept provide information about the amount of error in the self-monitoring system (e.g. a slope and/or y-intercept that differ from 1.0 and 0.0, respectively, are generally believed to indicate that there is major error in the self-monitoring results) may be incorrect under certain conditions. One of these conditions is when there are outliers or extreme values in the data. An outlier is usually defined as a measurement that deviates from the average blood glucose measurement by more than three standard deviations, and it may reflect a truly anomalous measurement or it may be the result of a problem in collecting or recording the data. Outliers in the data will result in poor estimates of the slope and y-intercept, and consequently will produce misleading information about the amount of error associated with the use of a self-monitoring device. The data in Table 4 illustrate how even one extreme value or outlier in the measurements can have a large effect on the values of the slope and y-intercept. The regression line for these data has a y-intercept of -3.2 and a slope of 2.0 ($y = 2.0x + -3.2$) and would suggest that there are serious discrepancies between the self-monitoring and laboratory values. Upon closer examination of the measurements, however, it is apparent that there is perfect agreement for all but one measurement (an outlier). While outliers have more pronounced effects when the

Table 3. Hypothetical glucose values (mmol l^{-1}) from a self-monitoring device and a laboratory method illustrating a regression line with a slope of 1.0 and a *y*-intercept of 0.0

Laboratory reference value (mmol l^{-1})	Self-monitoring value (mmol l^{-1})	% Deviation	Error grid analysis zone	Clinical significance
3.3	0.5	84.8	A	Appropriate clinical decision
3.3	6.1	84.8	D	Dangerous failure to treat
4.2	1.4	66.7	B	Benign error
4.2	7.0	66.7	B	Benign error
4.4	1.6	63.6	B	Benign error
4.4	7.2	63.6	B	Benign error
5.6	2.8	50.0	B	Benign error
5.6	8.4	50.0	B	Benign error
11.1	8.3	25.2	B	Benign error
11.1	13.9	25.2	B	Benign error

Table 4. Hypothetical glucose values (mmol l^{-1}) from a self-monitoring device and a laboratory method illustrating the influence of an outlier

Laboratory reference value (mmol l^{-1})	Self-monitoring value (mmol l^{-1})	% Deviation	Error grid analysis zone	Clinical significance
1.7	1.7	0.0	A	Clinically accurate
2.2	2.2	0.0	A	Clinically accurate
2.8	2.8	0.0	A	Clinically accurate
3.3	3.3	0.0	A	Clinically accurate
3.9	3.9	0.0	A	Clinically accurate
4.4	4.4	0.0	A	Clinically accurate
5.0	5.0	0.0	A	Clinically accurate
5.5	5.5	0.0	A	Clinically accurate
6.1	6.1	0.0	A	Clinically accurate
6.7	16.7	149.2	C	Unnecessary correction

sample of measurements is small, outliers in large data sets can also affect the values of the slope and *y*-intercept, thus leading to misleading conclusions about the errors associated with a self-monitoring device. This issue is relevant to investigators examining the performance of self-monitoring systems both within the laboratory and in clinical practice.

A second condition that can seriously affect the estimates of the slope and *y*-intercept is a nonlinear relationship between the laboratory and self-monitoring measurements. If a curvilinear relationship exists between these measurements the least squares estimates of the slope and *y*-intercept will be distorted and consequently the estimates of the amount of error associated with the self-monitoring results will be inaccurate.

Conclusions

The Pearson product–moment correlation and linear regression technique are two of the most widely known and used statistical techniques, and are frequently used to evaluate the quality of the data obtained from self-monitoring systems. This paper has presented examples illustrating the limitations of these techniques and has attempted to dispel some of the incorrect notions concerning these statistics. In particular, it has been shown that a Pearson *r* of 1.00 or a regression equation with a slope of 1.0 and a *y*-intercept of 0.0 does not necessarily mean that the self-monitoring measurements are without error. In addition, this paper has shown that a Pearson *r* close to 0.0 or a regression line with a slope and *y*-intercept that differ significantly from 1.0 and 0.0, respectively, does not necessarily mean that the self-monitoring measurements deviate greatly from the true values.

In view of the limitations of the Pearson *r*, this statistic should not be used to evaluate either the amount of error associated with a self-monitoring system or the clinical significance of the errors. In addition, it is argued that linear regression should not be used to provide a quantitative estimate of the amount of error associated with a self-monitoring system if any of the following conditions are present:

1. outliers in the data;

2. nonlinear relationship between the self-monitoring and laboratory measurements;
3. a restricted range in the blood glucose measurements.

These conditions will lead to distorted estimates of the slope and y-intercept and misleading descriptions of the amount of error associated with the self-monitoring results. If these conditions are not present the slope, y-intercept, and standard error of estimate can be used to quantify the amount of proportional, constant, and random error, respectively. This information by itself, however, cannot be used to evaluate the clinical significance of the errors in the self-monitoring system. As has been shown throughout this paper, the meaning and clinical significance of the self-monitoring errors can only be determined by examining the magnitude of the errors in relation to the actual blood glucose values. Currently, the only technique that exists that takes into account these factors is error grid analysis,[19–21] and therefore, we would strongly recommend the use of this technique to determine the clinical significance of any errors in self-monitoring results.

Acknowledgements

This study was supported by National Institutes of Health Grant 2P60-DK-20572, National Institutes of Diabetes, Digestive and Kidney Diseases. We wish to acknowledge the assistance provided by V. Wilgus, L. Brooks and J. Rosan in preparing the manuscript.

References

1. Consensus Development Panel. Consensus statement on self-monitoring of blood glucose. *Diabetes Care* 1987; **10:** 95–99.
2. Averhahn C, Bergman M, Kumar SR, Morgan J. Reagent strip performance as evaluated by a meter. *The Diabetes Educator* 1985; **11:** 41–43.
3. Azis S, Hsiang Y. Comparative study of home blood glucose monitoring devices: Visidex, Chemstrip bG, Glucometer, and Accu-Chek bG. *Diabetes Care* 1983; **6:** 529–532.
4. Birch K, Hildebrandt P, Marshall MO, Sestoft L. Self-monitoring of blood glucose without a meter. *Diabetes Care* 1981; **4:** 414–416.
5. Canfield ME, Kemp SF, Hoff C. New blood glucose strip: Comparison of Visidex, Visidex II, and Chemstrip bG with glucose analyzer determination of blood glucose. *Diabetes Care* 1985; **8:** 77–82.
6. Clarson C, Daneman D, Frank M, Link J, Perlman K, Ehrlich RM. How accurate are children with diabetes at reading chemstrip bG? *Diabetes Care* 1985; **8:** 354–358.
7. Clements RS, Keane NA, Kirk KA, Boshell BR. Comparison of various methods for rapid glucose estimation. *Diabetes Care* 1981; **4:** 392–395.
8. Gifford-Jorgensen RA, Borchert J, Hassanein R, Tilzer L, Eaks GA, Moore WV. Comparison of five glucose meters for self-monitoring of blood glucose by diabetic patients. *Diabetes Care* 1986; **9:** 70–76.
9. Godine JE, Hurxthal K, Nathan DM. Bedside capillary glucose measurement by staff nurses in a general hospital. *Am J Med* 1986; **80:** 803–806.
10. Kubilis P, Rosenbloom AL, Lezotte D, *et al.* Comparison of blood glucose testing using reagent strips with and without a meter (Chemstrips bG and Dextrostix/Dextrometer). *Diabetes Care* 1981; **4:** 417–419.
11. Laus VG, Dietz MA, Levy RP. Potential pitfalls in the use of glucoscan and glucoscan II meters for self-monitoring of blood glucose. *Diabetes Care* 1984; **7:** 590–594.
12. Marquette GP, Dillard T, Bietla S, Nie Byl JR. The accuracy of visual and meter determinations of blood glucose with the use of Chemstrip bG. *Am J Obstet Gynecol* 1985; **153:** 883–884.
13. Most RS, Gross AM, Davidson PC, Richardson, P. The accuracy of glucose monitoring by diabetic individuals in their home setting. *The Diabetes Educator* 1986; **12:** 24–27.
14. Nelson JD, Woelk MA, Sheps, S. Self glucose monitoring: A comparison of the Glucometer, Glucoscan, and Hypocount B. *Diabetes Care* 1983; **6:** 262–267.
15. Schade DS, Eaton RP, Mitchell WJ, Friedman, NM. Intravenous home blood glucose monitoring. *Diabetes Care* 1981; **4:** 420–423.
16. Schiffrin A, Desrosiers M, Belmonte M. Evaluation of two methods of self blood glucose monitoring by trained insulin-dependent diabetic adolescents outside the hospital. *Diabetes Care* 1983; **6:** 166–169.
17. Wing RR, Epstein LH, Nowalk MP, Scott N, Koeske, R. Compliance to self-monitoring of blood glucose: A marked-item technique compared with self-report. *Diabetes Care* 1985; **8:** 456–460.
18. Westgard JO, Hunt, MR. Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 1973; **19:** 49–57.
19. Cox D, Clarke W, Gonder-Frederick L, *et al.* Accuracy of perceiving blood glucose in insulin-dependent diabetes mellitus. *Diabetes Care* 1985; **8:** 529–536.
20. Pohl S, Gonder-Frederick L, Cox D, Evans W. Self-measurement of blood glucose concentration: clinical significance of patient-generated measurements. *Diabetes Care* 1985; **8:** 617–619.
21. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 1987; **10:** 622–628.
22. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974; **20:** 825–833.
23. Robinson WS. The statistical measurement of agreement. *Am Sociol Rev* 1957; **22:** 17–25.