Clin Oral Invest (2005) 9: 65–69
DOI 10.1007/s00784-004-0290-3

ORIGINAL ARTICLE

Hans-Peter Müller · Ebrahim Behbehani

Methods for measuring agreement: glucose levels in gingival crevice blood

Received: 11 August 2003 / Accepted: 4 October 2004 / Published online: 6 January 2005

Abstract The aim of the present study was to compare conclusions drawn by two different methods for comparison of blood glucose determination in capillary fingerstick blood (CFB) and gingival crevice blood (GCB). Glucose levels in CFB and GCB oozing from the gingiva after periodontal probing were measured in 31 patients with gingivitis or periodontitis using a novel, very sensitive self-monitoring device (Freestyle, TheraSense Inc.) developed for off-finger tip glucose testing. Correlation analysis revealed that measurements of glucose levels in CFB from left and right finger tips were highly correlated pointing to excellent performance of the device, whereas CFB and GCB measurements were moderately, but highly significantly, correlated. A thorough analysis of agreement revealed, on the other hand, questionable performance of the device for screening hypoglycaemic patients. The mean difference of measurements in CFB samples was +3.2±12.7 mg/dl. The 95% limits of agreement were −21.7 and +28.2. The mean difference of glucose determination in CFB and GCB samples was −22.0±26.6 mg/dl, and limits of agreement were −74.4 and +30.1. By plotting differences on means of measurements and doing linear regression analysis no systematic trend of change in differences with increasing mean of measurements was ascertained. Analysis of agreement revealed that performance of the Freestyle measuring device yielded considerably large limits of agreement, and gingival crevice blood cannot be recommended for measuring blood glucose levels.

Keywords Coefficient of agreement · Correlation · Method comparison · Blood glucose · Periodontal disease

Introduction

Clinical diagnosis in medicine must be based on established techniques. The method with the highest level of diagnostic accuracy is known as the 'gold standard'. Its application can be extremely difficult without exerting adverse effects on the patient. An example may be quite invasive, histological evidence from biopsy. Therefore, the gold standard is frequently excluded from the routine armamentarium of the clinician. Instead, surrogate methods are developed which are easier to employ, cheaper, and/or exert less adverse effects. The new 'standard method' then has to prove sufficient agreement with the gold standard. As a matter of fact, the standard method cannot be measured without error [7, 8]. That is why some lack of agreement between different methods of measurement has to be expected. Concordance of diagnostic findings documented with different methods is frequently presented as the result of correlation and/or regression analysis. The intraclass correlation coefficient as a measure of reliability [6] is often used to document appropriateness of a new diagnostic method. However, what actually matters is the *amount of disagreement*. If a new method is to be introduced we want to know by how much it is likely to differ from the old one. The amount of disagreement that can be tolerated in a clinical setting depends entirely on the parameter in question. In fact, statistical methods employing even hypothesis testing cannot answer such a question [3].

The aim of the present study is to compare the conclusions drawn following the still widely used approach of correlation/regression analysis of comparative data of diagnostic methods, and more appropriate measures of calculating limits of agreement. Blood glucose levels as measured in capillary fingerstick (standard method) and gingival crevice blood in subjects with gingivitis or periodontitis were used for this comparison. A novel, very sensitive monitoring device requiring an amount of only 0.3 μl blood was used, which may allow less painful off-finger tip testing. Based upon recent reports on the feasibility of available glucometer devices for glucose determination in gingival crevice blood of periodontitis patients [1, 14], the original objective was to test the more sensitive self-monitoring glucometer in this regard in subjects with mild periodontal disease. The results of this study have been published elsewhere [13].

H.-P. Müller (✉) · E. Behbehani
Faculty of Dentistry, Kuwait University,
P.O. Box 24923, 13110 Safat, Kuwait
e-mail: hp.muller@hsc.edu.kw

Materials and methods

A total of 46 patients participated in the clinical study. All of them were seeking dental check-up or treatment at Kuwait University Dental Clinic. Twenty-four patients had plaque-induced gingivitis, and 22 presented with increased probing depth and attachment loss, periodontitis. The usual exclusion criteria for blood glucose determination applied. After briefing on the procedures and potential risks and benefits, patients gave their written consent for participation. Seven patients (15%) were aware of suffering from diabetes, 3 had type 1, and 4 type 2 diabetes mellitus.

Periodontal examination included measurement of probing depth, attachment level, and bleeding on probing. A site with more profuse bleeding was chosen for collecting the gingival crevice blood (GCB) sample. The area was isolated with cotton rolls to prevent saliva contamination and dried with compressed air. Probing was repeated until a sufficient amount of blood appeared in the gingival crevice. The FreeStyle Blood Glucose Monitoring System (TheraSense Inc., Alameda, CA, USA) was used according to the manufacturer's recommendations. Unlike common amperometric strips, the FreeStyle strip uses an osmium-based mediator that reacts at a very low electrochemical potential. This coulometric measurement method is unaffected by common interfering substances such as uric acid, aspirin, or acetaminophen, which can react at very low potential. The device requires only a droplet of 0.3 μl for accurate determination of blood glucose and is particularly recommended for off-finger glucose testing [5]. Immediately before measuring glucose levels in GCB, a capillary fingerstick blood (CFB) sample was drawn from the right index finger using a disposable sterile lancet. To determine reliability of the glucometer device, 17 patients provided another sample from the left index finger.

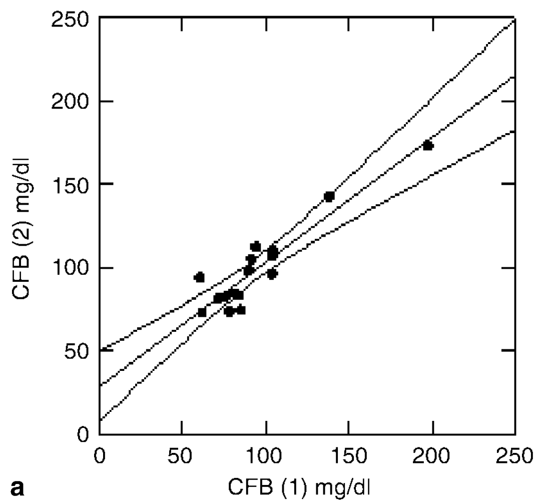
The statistical analysis of the blood glucose measurement pairs in each participant was done in two ways. First, the correlation was assessed by calculating Pearson's *r*. Linear regression analysis was employed and the intercept and regression coefficient were tested for 'significance'. Thereafter, a thorough analysis of agreement between the two methods was done according to Bland and Altman [2, 3]. This included analysis of data pairs, calculation of bias, and limits of agreement, as well as graphical display. If differences are assumed to be normally distributed, they are found with 95% confidence within limits defined as mean difference ±1.96 times the standard deviation (SD) of differences (coefficient of agreement), the limits of agreement. If limits of agreement are assumed to follow a normal distribution, and unless *n* is small, the standard error (SE) of the upper and lower limit of agreement is 1.71 times the standard error of the mean difference. By considering the *t* -distribution and *n* -1 degrees of freedom (df), the 95% confidence interval (CI) can be calculated as *t* standard errors either side of the observed value. Graphical display included scatter plots of one measurement against the other, and differences against means.

Results

Among 46 patients, blood glucose in GCB could be determined in only 31 cases. Readings ranged between 21 and 180 mg/dl. In the remaining 15 cases, samples were too small for measurement. In one patient with type I diabetes mellitus, CFB sampling revealed an error message (>500 mg/dl). Readings ranged between 25 and 207 mg/dl. A total of 31 comparisons of CFB and GCB measurements were possible.

Correlation/regression analysis

The 17 paired CFB samples from right and left index fingers revealed a high correlation coefficient of $r=0.93$, which was highly significant ($P<0.001$). A moderate albeit highly



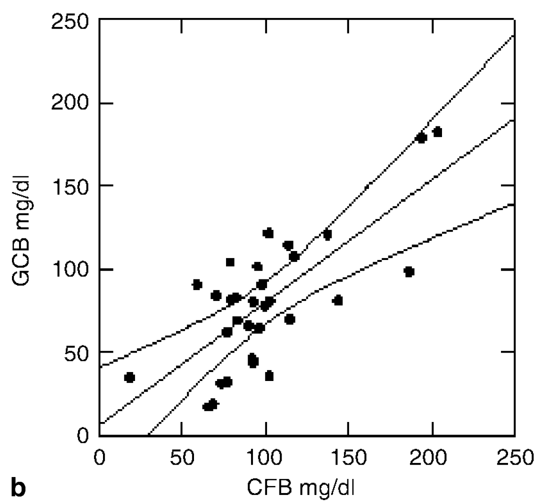


Fig. 1 a Linear regression (95% confidence bands) of glucose levels measured in the second capillary fingerstick blood sample (CFB2) on those measured in the first sample (CFB 1). **b** Linear regression of gingival crevice blood (GCB) sample measurements on CFB sample readings.

Table 1 Analysis of repeatability/agreement of measurements of blood glucose levels (mg/dl) in capillary fingerstick blood (*CFB*) and gingival crevice blood (*GCB*) samples. *SD* standard deviation, *CI* confidence interval

	CFB (1), CFB (2) (*n* =17)	CFB, GCB (*n* =31)
Minimum difference	−33	−84
Maximum difference	+32	+23
Mean difference ± SD	3.2±12.7	−22.0±26.6
Coefficient$_{95\%}$ of agreement	24.9	52.1
Limits$_{95\%}$ of agreement	−21.7, +28.2	−74.4, +30.1
95% CI of lower limit	−32.9, −10.5	−90.8, −57.5
95% CI of upper limit	17.0, 39.3	13.4, 46.8

significant correlation was seen between GCB and CFB readings (r =0.75, P <0.001). Regression of GCB on CFB revealed an intercept of 3.9±12.8 mg/dl (not significant) and a regression coefficient of 0.74±0.12 (P <0.001). Correlation between glucose levels in CFB samples, as well as CFB and GCB sample readings are displayed in Fig. 1.

Analysis of agreement

Table 1 presents the results of a thorough analysis of agreement when comparing blood glucose measurements. The mean difference of measurements in CFB samples was +3.2±12.7 mg/dl. Therefore, the 95% coefficient of agreement was 24.9 mg/dl (1.96×SD of differences). If the differences were normally distributed, 95% of the differences were expected to lie within the interval between −21.7 and +28.2 mg/dl. Considering 16 df and t =2.12, the

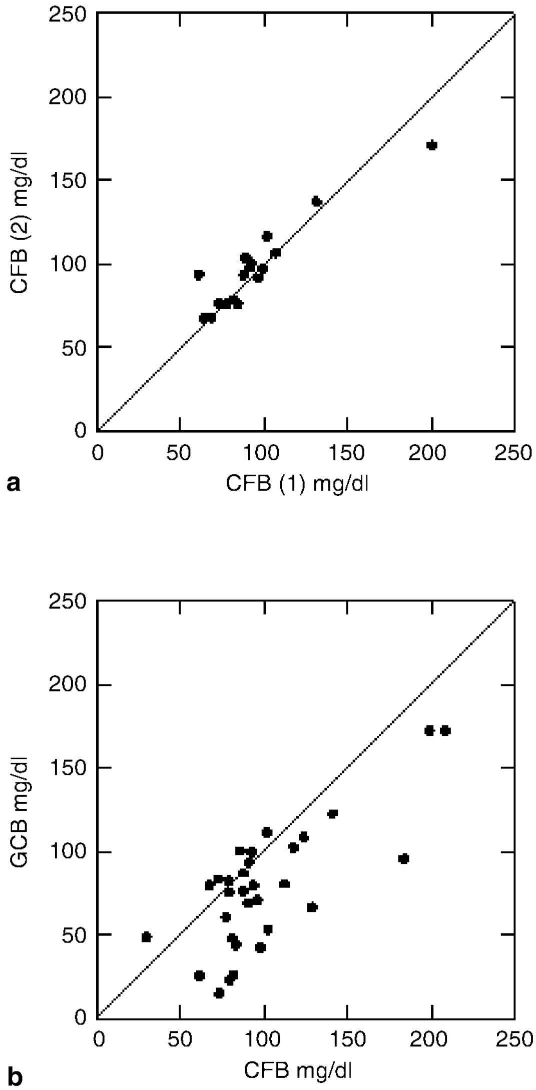


Fig. 2 a CFB (1) and CFB (2), and **b** CFB and GCB readings with lines of equality.

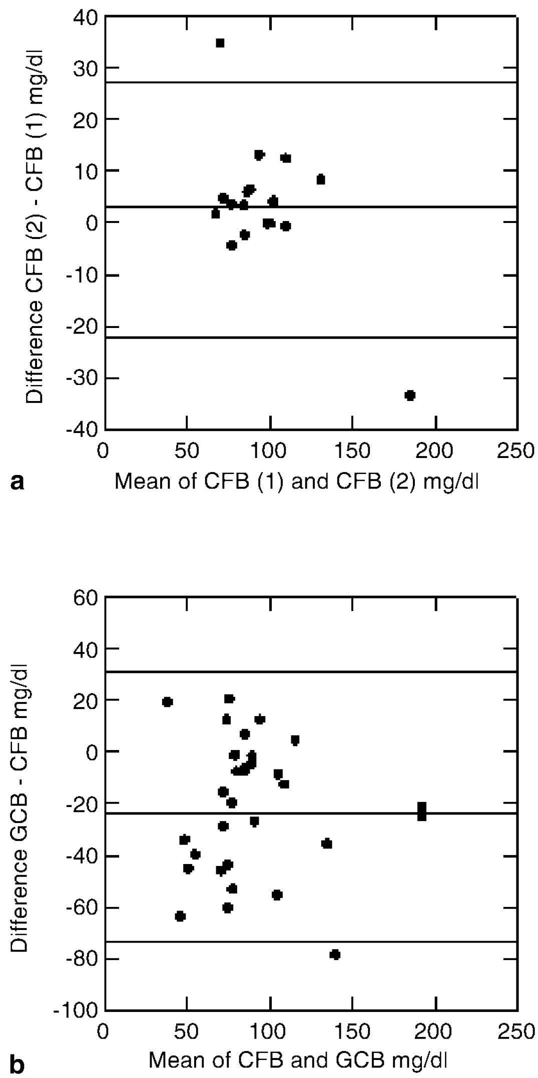


Fig. 3 a Plots of means of measurements in CFB (1) and CFB (2), as well as **b** CFB and GCB against respective differences. The mean differences as well as upper and lower 95% limits of agreement (mean ± 1.96×SD) are given.

precision of these estimates is given by the 95% CI of the lower (−32.9 to −10.5) and the upper limit of agreement (17.0 to 39.3). The mean difference of glucose in CFB and GCB samples was −22.0±26.6 mg/dl with limits of agreement of −74.4 and +30.1 mg/dl (Table 1).

Instead of regressing the second CFB, or GCB, readings on CFB measurements (erroneously either assuming dependence of measurements or desiring prediction of a second measurement), in Fig. 2 the *line of equality* is drawn when plotting one measurement against the other. The plot does not provide evidence of a systematic bias in CFB comparisons; however, glucose levels in GCB were generally lower. In Fig. 3, differences between measurements are plotted against their respective averages. In addition, lines mark the mean difference as well as limits of agreement. No proportional error (systematic trend of change in differences with higher mean values of measurements) in either comparison was ascertained. To further substantiate this, differences (D) were regressed on the means (M) of measurements: D =−20.2−0.02 M mg/dl. Both regression parameters were not ‘significant’ (P values of 0.139 and 0.884 respectively). Finally, absolute values of the residuals from the model were regressed on M (P =0.250).

Discussion

The American Academy of Periodontology recently stated in a position paper on diabetes and periodontal disease: “Glucometers are commonly used by diabetic patients for home monitoring of their blood glucose levels using a single drop of blood from a fingerstick. This procedure is of interest to the dental practitioner since it is simple, relatively inexpensive, and of sufficient accuracy to serve as an in-office screening device for patients suspected to have diabetes, and to monitor blood sugar levels of known diabetics” [12]. The Freestyle glucometer is a novel, very sensitive device which had explicitly been developed for off-finger tip glucose testing [11], in particular the forearm. It requires a tiny volume of blood of about 0.3 µl. The performance is impaired in situations of rapid decrease of glucose concentrations in the state of hypoglycaemia [11], and immediately postprandial [9]. In the present study, glucose concentrations were directly measured with the Freestyle glucometer in gingival blood oozing from the sulcus after routine periodontal probing. Possible advantages of measuring glucose levels directly in GCB may be lack of any pain, and easy and safe collection of the sample [1]. With r =0.93, the correlation between subsequent measurements in right and left finger prick samples was reasonably high, and highly significant. In particular, no systematic bias of the measurements was ascertained. Analysis of agreement revealed, on the other hand, that 95% of differences could be found within an interval of ±25 mg/dl about the mean difference of about 3 mg/dl. When patients are screened for hyperglycaemia, this agreement may be considered sufficient; it is clearly too large, however, for screening hypoglycaemic patients [10].

For analytical reasons only, correlation/regression was also conducted when comparing blood glucose in CFB and GCB samples. In that case, correlation was moderate (r =0.75). The intercept was insignificant, while the regression coefficient of 0.74 was highly significant. Admittedly, the conclusions drawn by looking at the results of correlation/regression analysis of the data can be similar to those discussed next. As a matter of fact, however, serious misconceptions are merely misleading when applying this approach of analysis for measurement comparison. First, correlation depends on the range of the variables. Obviously, correlation measures association, not agreement. Therefore, any overall summary measure, such as the correlation coefficient, does not help a clinician interpret a measurement. Regression analysis attempts to predict an observed measurement by another observed measurement. It suggests that one measurement can be modelled by another, which is not the case in measurement comparison. There is no independent measurement since both (surrogate) measurements definitively deviate from the true value. The key to method comparison is quantification of disagreement of the measurements [2], not of residuals. In particular, any hypothesis testing in measurement comparison is misleading.

As can be seen in Fig. 2b, glucose measurements in GCB were generally lower than those in finger prick samples. This might partly be explained by contamination of crevice blood with gingival exudate, in particular in tiny samples [13]. In addition, the large 95% coefficient of agreement of more than 50 mg/dl renders measurements of blood glucose in gingival blood in general not comparable with those in fingerstick blood. Because of the small sample size, the precision of the limits of agreement, given by their 95% confidence intervals, was rather low. They further underscore, however, that the degree of agreement was not acceptable. Plotting differences between measurements against their means (Fig. 3) enables checking important assumptions, namely that mean and SD of the differences were constant throughout the range of measurement, and that differences follow approximately Normal distribution. Also, disagreement is explicitly displayed, which is not so apparent in Fig. 2. Extreme observations (for example in Fig. 3b, there was a difference of −84 mg/dl in a pair of measurements) can easily be identified. The scatter plot also indicates that there was no systematic trend of increasing (or decreasing) differences with higher mean values. Linear regression analysis of differences on means of measurements is helpful to further substantiate this [3]. And finally, if necessary, the scatter of residuals from the model can be modelled as a function of the size of the measurement.

Another method of graphical display of blood glucose measurement comparison, the error grid analysis [4], is constructed upon the assumption that glucose levels should be kept within an ideal range of 70–180 mg/dl. Readings of a glucose monitor are compared to a laboratory method on a scatter plot. The graph is then divided into zones that represent the error as it would relate to therapy, e.g. zone A: clinically accurate, within ±20% of laboratory; zone B: benign error, greater than ±20% of laboratory; zones C, D, and E: errors leading to unnecessary corrective treatment, potentially dangerous failure to detect hypo- or hyperglycaemia, or erroneous treatment of hypo- or hyperglycaemia, respectively. While this approach may be appealing to some clinicians, the original introductory paper itself contained a fundamental error in constructing the upper A-limit, due to misconception of regression [15]. One should be aware that error grid plots published in the past may contain an erroneous upper A-line, and conclusions regarding agreement might have been too optimistic.

In the present study, a thorough analysis of agreement revealed that performance of the Freestyle measuring device itself yielded considerably large limits of agreement, which might especially be of concern in the state of hypoglycaemia. When comparing glucose measurements in GCB with those in CFB samples, no systematic trend of change of variation of the differences with higher measurements was noted. However, generally lower readings in gingival crevice blood pointed to considerable contamination with gingival exudate. A large coefficient of agreement must lead to the conclusion that such differences would be unacceptable for clinical purposes. Therefore, screening for elevated blood glucose levels should not be performed in gingival crevice blood oozing from the sulcus after routine periodontal probing.

References

1. Beikler T, Kuczek A, Petersilka G, Flemmig TF (2002) In-dental-office screening for diabetes mellitus using gingival crevicular blood. J Clin Periodontol 29:216–218
2. Bland JM, Altman DG (1986) Comparing methods for assessing agreement between two methods of clinical measurement. Lancet i:307–310
3. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160
4. Clarke WL, Cox D, Gender-Frederick LA, Cater W, Pohl SL (1987) Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 10:622–628
5. Feldman B, McGarraugh G, Heller A, Bohannon N, Skyler J, DeLeeuw E, Clarke D (2000) FreeStyle: a small-volume electrochemical glucose sensor for home blood glucose testing. Diabetes Technol Ther 2:221–229
6. Fleiss JL (1986) Reliability of measurement. In: Fleiss JL (ed) The design and analysis of clinical experiments. Wiley, New York, pp 1–32
7. Koran LM (1975) The reliability of clinical methods, data and judgements (first of two parts). N Engl J Med 293:642–646
8. Koran LM (1975) The reliability of clinical methods, data and judgements (second of two parts). N Engl J Med 293:695–701
9. Lee DM, Weinert SE, Miller EE (2002) A study of forearm versus finger stick glucose monitoring. Diabetes Technol Ther 4:13–23
10. McGarraugh G, Putz J (2002) FreeStyle(tm) accuracy at low glucose concentrations.http://www.therasense.com/freestyle-meter/pdf/ART02017_rev_A_REF.pdf
11. McGarraugh G, Price D, Schwartz S, Weinstein R (2001) Physiological influences on off-finger glucose testing. Diabetes Technol Ther 3:367–376
12. Mealey B (2000) Diabetes and periodontal disease. J Periodontol 71:664–678
13. Müller HP, Behbehani E (2004) Screening of elevated glucose levels in gingival crevice blood using a novel, sensitive self-monitoring device. Med Princ Pract 13:361–365
14. Parker RC, Rapley JW, Isley W, Spencer P, Killoy WJ (1993) Gingival crevicular blood for assessment of blood glucose in diabetic patients. J Periodontol 64:666–672
15. Stöckl D, Dewitte K, Fierens C, Thienpont LM (2000) Evaluating clinical accuracy of systems for self-monitoring of blood glucose by error grid analysis: comment on constructing the “upper A-line”. Diabetes Care 23:1711