

Measurement of blood glucose: comparison between different types of specimens

Bendix Carstensen¹, J Lindström², J Sundvall³, K Borch-Johnsen¹, J Tuomilehto^{2,4} and the DPS Study Group

¹Steno Diabetes Center, Niels Steensens Vej 2, DK-2820 Gentofte, Denmark; ²Diabetes Unit; ³Department of Health and Disability, National Public Health Institute; ⁴Department of Public Health, University of Helsinki, FI-00300 Helsinki, Finland
Corresponding author: Dr Bendix Carstensen. Email: bxc@steno.dk

Abstract

Aim: To provide conversion formulae between measurements based on different specimens in use in epidemiological studies and clinical practice, and to evaluate the relative precision for the different methods.

Background: The current guidelines emphasize the use of venous plasma for determining glucose concentration. Nevertheless, the World Health Organization (WHO) guidelines provide cut-off points for different specimens for the determination of the glucose concentration in circulating blood (venous plasma, whole blood, serum and capillary blood). There is a lack of data about the comparability between the values obtained by using different specimens.

Methods: Eleven different combinations of specimens and methods of measurement of blood glucose were used in 294 blood samples from 74 subjects. The methods were grouped by the specimen used for analysis (venous plasma, whole blood, serum and capillary blood).

Results: The result of the analysis is a set of linear equations allowing conversion of the result from one specimen or method to another. Furthermore, it was estimated how much of the variation for each method can be attributed to laboratory variance.

Conclusions: Measurements based on capillary blood had a very large variability compared with other methods.

Measurements based on venous whole blood tended to give results 0.5 mmol/L lower than other methods. Our data indicate that the current diagnostic cut-off points, as recommended by WHO for non-plasma specimens, are not fully compatible and may differ as much as 0.5 mmol/L between specimens.

Ann Clin Biochem 2008; 45: 140–148. DOI: 10.1258/acb.2007.006212

Introduction

In clinical diagnosis and monitoring of diabetes mellitus (DM), as well as in screening and in epidemiological studies of glucose tolerance, it is important to be able to produce accurate and reliable measurements of blood glucose levels. In practical clinical settings, it is of primary importance to have reproducible methods in order to make the surveillance of patients' circulating glucose levels as precise as possible. Hence, the primary concern will be the accuracy and reproducibility of the particular method in use at a particular clinic, i.e. internal validity.

Blood glucose is measured in many different ways (plasma/blood, capillary/venous, etc.) and the World Health Organization (WHO) provides diagnostic cut-off points for each of these, but no reference for the conversion factors is provided. The cut-off points for diagnosis of DM and glucose intolerance are therefore based on conversion factors for which there are very limited scientific data.

The present study aims at solving this lack of information by providing conversion formulae.

As a diagnosis of diabetes has major implications for the patient with respect to treatment, lifestyle, insurance, etc., it is important that the diagnosis based on specific values of the glucose concentration in peripheral circulation (typically cut-off points), is independent of the method of assay and laboratory used (external validity). The measured values should reflect those of the laboratories that formed the basis for the implemented clinical procedures. In epidemiological studies where populations are surveyed with respect to glucose tolerance, it is of further interest to know the relationship between methods used in different centres. Particularly in studies that comprise populations diverse in time or geography, one is likely to encounter a large number of methods used. Therefore, both systematic differences between methods (biases) and variations of the measurement methods with respect to repeatability,

reproducibility and possibly random interactions between individuals/samples and methods must be addressed. Both systematic and random differences between methods must in this context be considered at two levels:

- (1) Differences between methods based on different blood specimens (i.e. venous plasma/venous whole blood/venous serum/capillary blood);
- (2) Differences between methods based on the same specimen.

The latter differences both in terms of bias and variance must be expected to be the smaller of the two.

It is somewhat surprising that hardly any conversion factors between different types of specimens or methods exist in the contemporary medical literature.^{1,2} The aim of the present study has been to address these questions for 11 combinations of four specimens and eight methods for blood glucose measurement. However, the repeatability question has not been addressed separately, as the main purpose has been for applications in epidemiological studies. The repeatability must be addressed in specific studies for each laboratory. Further, it was the aim to provide guidelines for using the prediction from one method to another in practical epidemiological situations.

In addition, we aimed at comparing the WHO diagnostic criteria³ for diabetes for measurements based on venous plasma, venous whole blood and capillary blood with the conversion formulae we derive. Ideally, the cut-off points used in the WHO diagnostic criteria for the different types of specimens should coincide with the conversion formulae except for rounding errors.

Material and methods

Subjects

The study subjects were randomly selected from among the participants of the Finnish Diabetes Prevention Study (DPS)^{4,5} in five study centres, 11–19 per centre. The subjects were, according to inclusion criteria to the DPS, overweight (body mass index [BMI] > 25) and had impaired glucose tolerance (WHO 1999 criteria)³ at the beginning of the study. This selection of patients was deliberate in order to assure a broad range of glucose levels for the comparison study.

Blood sampling

The oral glucose tolerance test (OGTT) was administered in the morning after an overnight fast of 10–16 h. After collection of the fasting venous and capillary blood samples, the subjects drank 75 g of anhydrous glucose and 1.6 g citric acid in 250–300 mL of water in the course of 3 min. Capillary and venous blood samples were collected 30, 60 and 120 min after the test load from the antecubital vein with the subject in a sitting position. The skin was cleaned (with 70% ethanol etc.) and blood was drawn into vacuum tubes. If a tourniquet was used, it was opened

immediately after the needle had entered the vein. All the blood samples were drawn by experienced research nurses.

The capillary sample was obtained using an automatic lancet from the side of the fingertip, without squeezing the finger. The capillary sample was taken immediately after the venous samples.

Sample processing

All centres collected a reference sample in a citrate fluoride tube, which was mixed by turning it around gently and centrifuged 10 min (3400 rpm), within 15 min. Plasma was separated and dispensed into two tubes, of which one was mailed the same day to the central laboratory in an envelope and the other one was frozen and kept at -20°C and sent to the central laboratory packed in dry ice.

Venous serum sample was drawn into a plain vacuum tube and allowed to clot at room temperature for 30–60 min and was centrifuged, after which the serum was separated. The serum was kept in the refrigerator, and the glucose concentration was measured within 3 h.

Glucose measurement methods

The 11 methods fall in four classes based on the material (specimen) used for the measurements as shown in Table 1. The names refer to a combination of the method and the specimen.

Table 1 The 11 measurement methods compared, grouped by specimen. The names refer to those used in labelling plots and tables

Specimen	Description
Venous plasma:	
n.plas1	The National Public Health Institute (NPHI) central laboratory using an enzymatic method (glucose dehydrogenase, Roche Unimate 7, Cobas Mira Plus, Hoffman-La Roche, Basel, Switzerland)
n.plas2	The same method as n.plas1, but using the frozen and thawed samples
h.plas	Plasma glucose, using the Hemocue B-plasma glucose analyser (Hemocue AB). These plasma values are automatically derived from the measurements on whole blood by the Hemocue
m.plas	Plasma glucose, using the Bayer Technicon DAX48
k.plas	Plasma glucose using an enzymatic method (Granutest 100, Kone Pro, Kone instruments, Espoo, Finland)
Venous whole blood	
h.blood	Venous whole blood glucose, using the Hemocue B-plasma glucose analyser
Venous serum:	
h.serum	Serum glucose, using the Hemocue B-plasma glucose analyser. These values are automatically derived from the measurements on whole blood by the Hemocue
m.serum	Serum glucose, using the Bayer Technicon DAX48
s.serum	Serum glucose at the NPHI central laboratory, using an enzymatic method (glucose dehydrogenase)
Capillary blood:	
h.cap	Capillary glucose, using the Hemocue B-plasma glucose analyser
o.cap	Capillary glucose, using the OneTouch Basic (LifeScan)

Data

Data are hierarchically classified by centre, person, time and method, with five centres, 74 persons and four time-points (0, 30, 60 and 120 min) for each person. Table 2 shows which methods were applied to samples from each centre.

Statistical methods

The statistical standard for comparing methods of measurement are given in Bland and Altman (1986, 1999),^{6,7} where ‘limits of agreement’, and the so-called Bland-Altman plots are explained. However, in this study, the different samples from the same person are not independent, we have more than two methods to compare, and we want to devise an algorithm to convert between the methods. Hence, a statistical model extending the Bland-Altman approach was set up.

The data using the pairs of methods, actually directly compared, are shown in Figure 1. The assumption about constant difference (which underlies ‘limits of agreement’) was clearly not tenable, so a model with a linear relationship between methods was chosen.

The model was fitted to both the original and log-transformed data, to see whether the log-transformation produced a substantially better fit. We found no indication of non-linear relationships or increasing variance by glucose level that would have made the log-transformation preferable (data not shown). On the other hand, data were reasonably compatible with modelling on a log-scale; there was very little power in the study to discriminate between the two approaches. The actual predictions between measurement methods obtained from using the log-transformed data would have been virtually the same as those we got from using the untransformed data (data not shown).

We have therefore chosen to present the conversion formulae (Tables 3 and 4) for data on the original glucose-concentration scale.

Table 2 Types of blood specimens for the measurement of blood glucose and allocation

Method	Centre					Σ
	1	2	3	4	5	
n.plas1	76	40	67	46	60	289
n.plas2	76	44	67	46	60	293
h.plas	76			46		122
m.plas	76					76
k.plas		44				44
h.blood	76					76
h.serum	76					76
m.serum	76					76
s.serum			67			67
h.cap	76			47		123
o.cap					60	60
Measurements (n)	608	128	201	185	180	1302
Samples (n)	76	44	67	47	60	294
Individuals (n)	19	11	17	12	15	74

Number of samples by centre. Note that a few samples were lost in some of the centres, so not all individuals were measured on all methods, and not all individuals contribute four samples

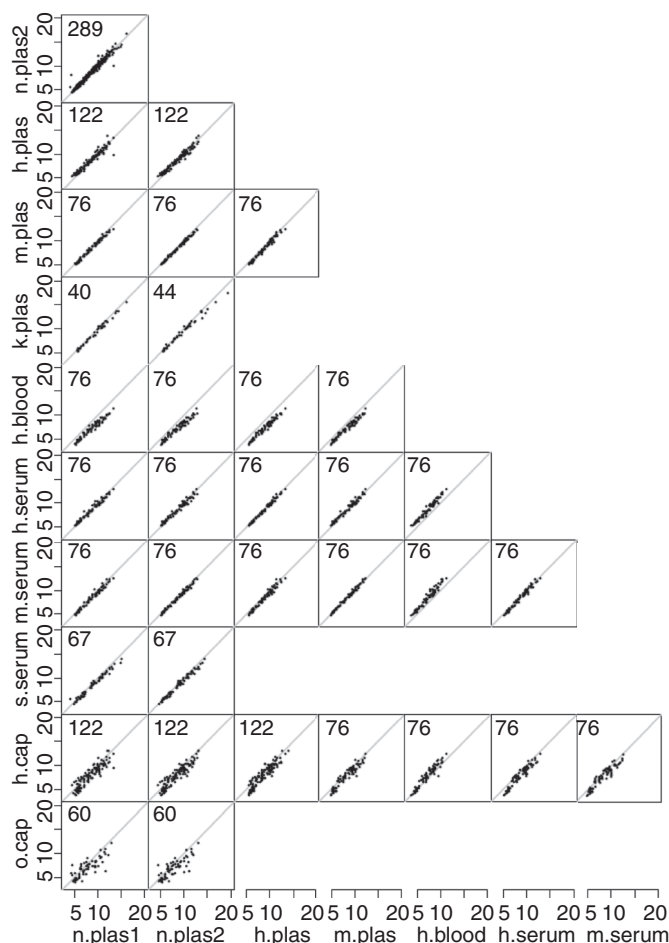


Figure 1 Original data (blood glucose in mmol/L) for the pairs of methods that have been applied to the same blood samples are shown. Number of points is given in the upper left corner of each panel

Model

The purpose of the statistical model is

- (1) To estimate simple linear conversion formulae between methods and
- (2) To give prediction limits for them.

The model used is a slight variant of the one described by Carstensen in 2004.⁸ We assume a simple linear relationship between methods, and that the relationship is the same at time-points 0, 30, 60 and 120 min. Statistically, this is expressed as a linear relation of each measurement y to an unknown ‘true’ value of the glucose concentration, μ . Formally, the measurement y_{mit} with method m on individual i at time t is modelled as:

$$y_{mit} = \alpha_m + \beta_m \mu_{it} + c_{mi} + e_{mit} \tag{1}$$

Where c_{mi} and e_{mit} are error terms. The random individual \times method interaction, c_{mi} , and the combined measurement error and sample \times method interaction, e_{mit} were assumed independent and normally distributed with means 0 and variances τ_m^2 and σ_m^2 respectively.

Table 3 Table of conversion formulae with prediction standard deviations

Predicted	Predictor, x										
	n.plas1	n.plas2	h.plas	m.plas	k.plas	h.blood	h.serum	m.serum	s.serum	h.cap	o.cap
n.plas1	0.000 (0.378)	+ 1.000x (0.369)	+ 1.094x (0.449)	+ 1.043x (0.317)	+ 1.043x (0.355)	+ 1.144x (0.452)	+ 1.090x (0.392)	+ 1.027x (0.345)	+ 1.086x (0.365)	+ 1.087x (0.895)	+ 1.255 (1.300x)
n.plas2	0.015 (0.369)	+ 1.000x (0.318)	+ 1.094x (0.416)	+ 1.043x (0.272)	+ 1.043x (0.293)	+ 1.144x (0.415)	+ 1.090x (0.354)	+ 1.027x (0.305)	+ 1.085x (0.323)	+ 1.087x (0.878)	+ 1.240 (1.300x)
h.plas	0.866 (0.410)	+ 0.914x (0.380)	+ 1.000x (0.374)	+ 0.953x (0.330)	+ 0.953x (0.347)	+ 1.046x (0.463)	+ 0.996x (0.404)	+ 0.939x (0.356)	+ 0.992x (0.377)	+ 0.939x (0.900)	+ 1.188x (1.192)
m.plas	0.051 (0.359)	+ 0.959x (0.261)	+ 1.049x (0.346)	+ 1.000x (0.129)	+ 1.000x (0.194)	+ 1.097x (0.338)	+ 1.045x (0.270)	+ 0.985x (0.216)	+ 1.041x (0.230)	+ 1.042x (0.848)	+ 1.152 (1.074)
k.plas	0.019 (0.321)	+ 0.958x (0.281)	+ 1.049x (0.364)	+ 1.000x (0.194)	+ 1.000x (0.172)	+ 1.097x (0.359)	+ 1.044x (0.292)	+ 0.985x (0.241)	+ 1.040x (0.256)	+ 1.042x (0.856)	+ 1.183 (1.082)
h.blood	+ 0.874x (0.395)	+ 0.874x (0.363)	+ 0.956x (0.443)	+ 0.912x (0.309)	+ 0.912x (0.327)	+ 1.000x (0.308)	+ 0.952x (0.385)	+ 0.898x (0.337)	+ 0.949x (0.357)	+ 0.950x (0.891)	+ 1.136x (1.123)
h.serum	0.622 (0.360)	+ 0.918x (0.325)	+ 1.004x (0.406)	+ 0.957x (0.258)	+ 0.957x (0.280)	+ 1.050x (0.404)	+ 1.000x (0.220)	+ 0.943x (0.293)	+ 0.996x (0.311)	+ 0.998x (0.874)	+ 1.193x (1.103)
m.serum	+ 0.974x (0.335)	+ 0.974x (0.297)	+ 1.065x (0.380)	+ 1.016x (0.219)	+ 1.016x (0.245)	+ 1.114x (0.376)	+ 1.061x (0.311)	+ 1.000x (0.224)	+ 1.057x (0.277)	+ 1.058x (0.862)	+ 1.319 (1.265x)
s.serum	0.297 (0.336)	+ 0.921x (0.298)	+ 1.008x (0.381)	+ 0.961x (0.221)	+ 0.961x (0.246)	+ 1.054x (0.377)	+ 1.004x (0.312)	+ 0.946x (0.262)	+ 1.000x (0.247)	+ 1.002x (0.863)	+ 1.265x (1.090)
h.cap	0.038 (0.823)	+ 0.920x (0.808)	+ 1.007x (0.906)	+ 0.959x (0.814)	+ 0.960x (0.821)	+ 1.052x (0.938)	+ 1.002x (0.876)	+ 0.945x (0.815)	+ 0.998x (0.861)	+ 1.000x (1.108)	+ 1.116 (1.463)
o.cap	0.965 (0.867)	+ 0.769x (0.853)	+ 0.842x (0.953)	+ 0.803x (0.862)	+ 0.803x (0.869)	+ 0.880x (0.989)	+ 0.838x (0.925)	+ 0.790x (0.861)	+ 0.835x (0.911)	+ 0.837x (1.224)	+ 1.000x (1.314)

These are used to construct prediction intervals: 90% intervals using $\pm 1.645 \times \text{SD}$, 95% intervals using $\pm 1.960 \times \text{SD}$. The SDs on the diagonal are $\sqrt{2\sigma_m^2}$ for the measurement method, reflecting that predictions are for a repeat sample from the same individual measured by the same method, hence excluding the individual \times method variation

Table 4 Table of conversion formulae with prediction standard deviations

Predicted	Predictor, x			
	Venous plasma	Venous blood	Venous serum	Capillary blood
Venous plasma	0.000 + 1.000 x (0.425)	0.558 + 1.119 x (0.461)	-0.137 + 1.047 x (0.407)	0.102 + 1.066 x (0.908)
Venous blood	-0.498 + 0.894 x (0.412)	0.000 + 1.000 x (0.309)	-0.621 + 0.936 x (0.392)	-0.407 + 0.953 x (0.901)
Venous serum	0.131 + 0.955 x (0.389)	0.664 + 1.069 x (0.419)	0.000 + 1.000 x (0.338)	0.228 + 1.018 x (0.889)
Capillary blood	-0.096 + 0.938 x (0.851)	0.428 + 1.050 x (0.945)	-0.224 + 0.982 x (0.873)	0.000 + 1.000 x (1.125)

Based on the data for the four different specimens, ignoring methods for glucose determination, and excluding all measurements by o.cap. The standard deviations are used to construct prediction intervals: 90% intervals using $\pm 1.645 \times \text{SD}$, 95% intervals using $\pm 1.960 \times \text{SD}$. The SDs on the diagonal are $\sqrt{2\sigma_m^2}$ for the measurement method, reflecting that predictions are for a repeat sample from the same individual measured by the same method, hence excluding the individual \times method variation

Thus, τ_m represents the standard deviation (SD) of the errors that cannot be changed by improved laboratory practice, whereas σ_m represents measurement error that is attributable to the technical laboratory procedure. The size of these sources of error is allowed to vary by the method, both for the individual by methods interaction and for the pure measurement error.

The 'true' sample values, μ_{it} , are taken as parameters and not assumed to follow any distribution, because subjects chosen for method comparison studies cannot reasonably be assumed to be representative of any population, and least of all populations on which derived prediction rules are to be applied. Seen from a prediction point of view, the μ values must be regarded as nuisance parameters of no interest *per se*, but any distributional assumption about them would be arbitrary and would likely lead to biased results with respect to predictive power.

Estimation

The model used is an extension of a two-way analysis of variance (two-factor main effects model), where a special feature of the interaction is modelled in the mean value structure as a linear relationship between the methods. Carstensen⁸ has devised an algorithm to estimate in a slightly simpler situation than this. This method has been applied to this data-set for the estimation of parameters.

Model validation

The model used requires a number of assumptions to be met:

- (1) The relationships between methods are linear;
- (2) The residuals are homoschedastic, i.e. the size of the variance components does not vary with the level of blood glucose;
- (3) The residuals are normally distributed.

Assumptions 1 and 2 were checked by plotting residuals against predicted means, and assumption 3 was checked by normal probability plots of the residuals. These results showed that the stated assumptions were met.

Prediction between specimens and methods

Prediction of measurements by one specimen/method from results obtained by another was based on this model. Prediction of a measurement by method 1, y_1 , from a measurement by method 2, y_2 , under model (1) is:

$$y_1 = \alpha_1 + \beta_1 \frac{y_2 - \alpha_2}{\beta_2} = \left(\alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2} \right) + \frac{\beta_1}{\beta_2} y_2$$

The intercept and slope on the right hand side of this equation is reported in Tables 3 and 4. The prediction SD was computed as:

$$\text{SD}(\text{prediction}) = \sqrt{\tau_1^2 + \sigma_1^2 + \frac{\beta_1^2}{\beta_2^2} (\tau_2^2 + \sigma_2^2)}$$

taking both the variation in y_1 and y_2 into account. Prediction intervals based on these formulae will be the same for predicting y_1 from y_2 or *vice versa*. The statistical method used here thus avoids the problems arising from regressing measurements by one combination of specimen/method on another which inevitably produces a relationship different from that obtained by regression the other way round. Furthermore, this model allows prediction between any two of the 11 combinations of specimen/method, even if direct pair-wise comparisons only exist in the data for 34 out of the possible 55 pairs.

Prediction from one combination of specimen and method to itself is, in essence, a prediction of the range where one would expect to see a second measurement on the same specimen by the same method of the same sample, and hence does not involve a new realization of the individual \times method interaction, and so has a prediction SD of

$$\sqrt{2\sigma_m^2}$$

Grouping methods by type of specimen

In order to be able to compare methods based on the four specimens venous plasma, whole blood, serum and capillary blood, we estimated a model as (1) where methods were grouped by specimen, i.e. all measurements on venous

plasma were considered one method etc. In this analysis with only four methods, the estimated μ_s from the analysis of the 11 individual methods were fixed and the conversion parameters and error terms estimated for these. However, as the plasma-based methods were represented by 824 measurements, out of which many were replicates on the same samples, we should expect to see a very small individual \times method (=specimen) interaction in this analysis, biasing the prediction SDs slightly towards 0.

Results

The estimates of conversion formulae between the 11 specimens/methods are given in Table 4. From the last two rows/columns in the table it is clearly seen that the methods based on capillary blood are by far the most imprecise methods, as the SDs of the predictions are about three times the size of those of the other methods. Further, it is clear from the last row/column that the o.cap method is systematically wrong. Subsequent to the study, it was verified that the analyser used was old and maladjusted. Hence, for the rest of the analyses these measurements have been excluded. From Table 5, it is seen that the residual variation is larger than the individual \times method interaction, except for h.blood and h.serum. Comparisons between methods by specimen are given in Table 4, example conversions in Table 6, and a graphical version in Figure 2. Clearly, the methods based on capillary blood have the lowest precision: the prediction intervals are more than twice as wide as those for the other substances. The best agreement is between venous plasma and serum; in the range 3–15 mmol/L, the disagreement is smaller than 0.5 mmol/L with serum values being lower. For venous blood, the readings tend to be around 1.5 mmol/L smaller than those from plasma and serum. For conversions between these three methods, the prediction standard error is about 0.4 mmol/L.

On average there is quite good agreement between capillary blood and serum, with capillary measurements being about 0.5 mmol/L lower. However, the method based on

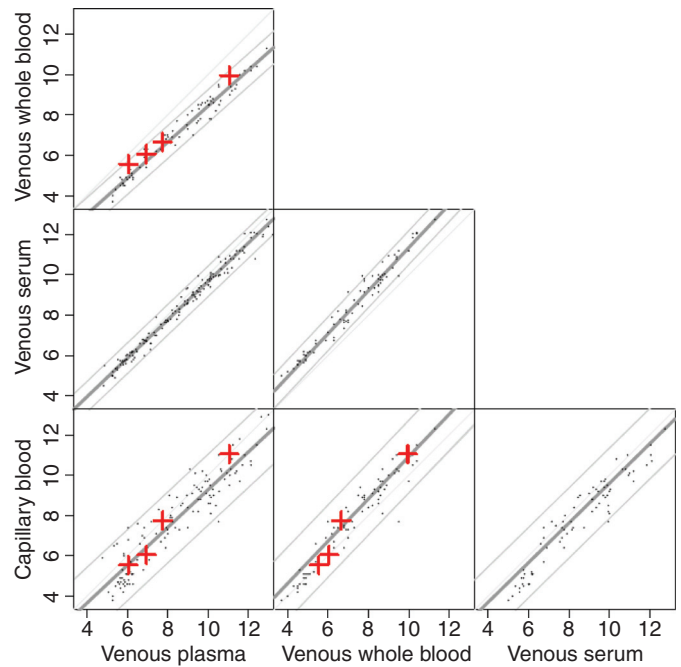


Figure 2 Graphical display of the prediction algorithms between the four specimens with 95% prediction limits as given in Table 4. The dots represent means over measurements by different methods based on the same substance. The crosses correspond to the diagnostic cut-off points laid down by the WHO consultation criteria

capillary blood is more imprecise, prediction standard errors between the other methods and capillary blood are all close to 1 mmol/L, i.e. any prediction interval will be of the order ± 2 mmol/L.

Finally, we applied our conversion factors compared with plasma to reproduce the table given in the report of the WHO consultation for different types of specimens (Table 7). Our data indicate that the cut-off points proposed by the WHO consultation were not in keeping with our results.

Discussion

The main aim of the present study was to provide conversion formulae between measurements of glucose based on different kinds of specimen. Table 4 provides this result. We have deliberately not addressed the issue of possibly differing conversions for samples drawn fasting, after a 2 h OGTT or just at a random time.

Other studies have investigated relationships for 2 h measurements only. Neely *et al.*¹ investigated relationships between measurements based on venous plasma, venous whole blood and capillary blood. They found the relationships:

$$\begin{aligned} \text{venous plasma} &= -0.08 + 1.10 \times \text{venous blood and} \\ \text{venous plasma} &= -0.06 + 1.10 \times \text{venous blood} \end{aligned}$$

for pregnant and non-pregnant individuals respectively. The relationship we found was:

$$\text{venous plasma} = 0.56 + 1.12 \times \text{venous blood}$$

(i.e. plasma values about 0.6 mmol/L higher throughout the range). Neely *et al.* found no variation with the haematocrit

Table 5 Table of the standard deviations of the variance components (mmol/L) from the analysis of 11 methods and from the analysis of the four specimens

	Individual \times method (τ_m)	Residual (σ_m)
Method		
n.plas1	0.093	0.268
n.plas2	0.074	0.225
h.plas	0.188	0.265
m.plas	0.069	0.091
k.plas	0.099	0.122
h.blood	0.227	0.218
h.serum	0.185	0.155
m.serum	0.094	0.159
s.serum	0.090	0.175
h.cap	0.303	0.783
o.cap	0.519	0.929
Specimen		
Venous plasma	0.005	0.306
Venous blood	0.180	0.219
Venous serum	0.030	0.239
Capillary blood	0.271	0.798

Table 6 Conversions with 95% prediction intervals from the formulae in Table 4

Measured value	Predicted value			
	Venous plasma	Venous blood	Venous serum	Capillary blood
Venous plasma				
5.0	5.0 (4.2; 5.8)	4.0 (3.2; 4.8)	4.9 (4.1; 5.7)	4.6 (2.9; 6.3)
6.0	6.0 (5.2; 6.8)	4.9 (4.1; 5.7)	5.9 (5.1; 6.6)	5.5 (3.9; 7.2)
7.0	7.0 (6.2; 7.8)	5.8 (5.0; 6.6)	6.8 (6.1; 7.6)	6.5 (4.8; 8.1)
8.0	8.0 (7.2; 8.8)	6.7 (5.8; 7.5)	7.8 (7.0; 8.5)	7.4 (5.7; 9.1)
9.0	9.0 (8.2; 9.8)	7.5 (6.7; 8.4)	8.7 (8.0; 9.5)	8.3 (6.7; 10.0)
10.0	10.0 (9.2; 10.8)	8.4 (7.6; 9.2)	9.7 (8.9; 10.4)	9.3 (7.6; 11.0)
Venous blood				
5.0	6.2 (5.2; 7.1)	5.0 (4.4; 5.6)	6.0 (5.2; 6.8)	5.7 (3.8; 7.5)
6.0	7.3 (6.4; 8.2)	6.0 (5.4; 6.6)	7.1 (6.3; 7.9)	6.7 (4.9; 8.6)
7.0	8.4 (7.5; 9.3)	7.0 (6.4; 7.6)	8.1 (7.3; 9.0)	7.8 (5.9; 9.6)
8.0	9.5 (8.6; 10.4)	8.0 (7.4; 8.6)	9.2 (8.4; 10.0)	8.8 (7.0; 10.7)
9.0	10.6 (9.7; 11.5)	9.0 (8.4; 9.6)	10.3 (9.5; 11.1)	9.9 (8.0; 11.7)
10.0	11.7 (10.8; 12.7)	10.0 (9.4; 10.6)	11.4 (10.5; 12.2)	10.9 (9.1; 12.8)
Venous serum				
5.0	5.1 (4.3; 5.9)	4.1 (3.3; 4.8)	5.0 (4.3; 5.7)	4.7 (3.0; 6.4)
6.0	6.1 (5.3; 6.9)	5.0 (4.2; 5.8)	6.0 (5.3; 6.7)	5.7 (4.0; 7.4)
7.0	7.2 (6.4; 8.0)	5.9 (5.2; 6.7)	7.0 (6.3; 7.7)	6.7 (4.9; 8.4)
8.0	8.2 (7.4; 9.0)	6.9 (6.1; 7.6)	8.0 (7.3; 8.7)	7.6 (5.9; 9.3)
9.0	9.3 (8.5; 10.1)	7.8 (7.0; 8.6)	9.0 (8.3; 9.7)	8.6 (6.9; 10.3)
10.0	10.3 (9.5; 11.1)	8.7 (8.0; 9.5)	10.0 (9.3; 10.7)	9.6 (7.9; 11.3)
Capillary blood				
5.0	5.4 (3.7; 7.2)	4.4 (2.6; 6.1)	5.3 (3.6; 7.1)	5.0 (2.8; 7.2)
6.0	6.5 (4.7; 8.3)	5.3 (3.5; 7.1)	6.3 (4.6; 8.1)	6.0 (3.8; 8.2)
7.0	7.6 (5.8; 9.3)	6.3 (4.5; 8.0)	7.4 (5.6; 9.1)	7.0 (4.8; 9.2)
8.0	8.6 (6.9; 10.4)	7.2 (5.5; 9.0)	8.4 (6.6; 10.1)	8.0 (5.8; 10.2)
9.0	9.7 (7.9; 11.5)	8.2 (6.4; 9.9)	9.4 (7.6; 11.1)	9.0 (6.8; 11.2)
10.0	10.8 (9.0; 12.5)	9.1 (7.4; 10.9)	10.4 (8.7; 12.2)	10.0 (7.8; 12.2)

For example, from a measured value of 7.0 mmol/L in a venous blood sample we would expect that the corresponding measured value in a venous plasma sample would be 8.4 mmol/L and that it with 95% probability would be between 7.5 and 9.3 mmol/L.

values, despite considerable differences between pregnant and non-pregnant individuals. Farrer *et al.*² found the relationship

$$\text{venous plasma} = -0.24 + 1.15 \times \text{venous blood}$$

also for 2 h postload measurements. However they found the ratio of measurements based on plasma to those based on whole blood to be increasing by haematocrit values.

Burrin and Alberti reported 14–16% higher glucose values in plasma compared with whole blood.⁹ This is a consequence of the higher water content for plasma (93%) compared with that of the erythrocyte (73%). Capillary blood is an intermediate mix of blood from arterioles, venules and capillaries. Thus, the difference in glucose concentrations between venous whole blood and capillary blood is small, but can be up to 8% higher in capillary blood after meals or glucose challenge.⁹ In addition, high haematocrit can result in additional bias.^{10,11}

These variations in the relationships are in this study embedded in the estimated variance components, and

incorporated in the prediction variation. It should be noted that the prediction variation is independent of the size of the study – the sources of variation (individual \times method and residual) are present in any new measurement by a particular method. The statistical uncertainty in the conversion formulae is negligible compared with these, and is therefore not reported here.

The variation between patients with respect to glucose tolerance is controlled for in the model, but the interaction between the glucose tolerance and the measurement method is taken account of in the variance components, and hence reflected in the prediction precision.

Therefore, the prediction variation estimated from this study is presumably larger than those that would be obtained if data were restricted to the 2 h values alone. However, this has not been possible to determine because we then would only have had a single measurement per person and specimen/method, and hence no way to estimate the variance components.

Table 7 Cut-off points (mmol/L) for diabetes and impaired glucose tolerance for venous plasma and for measurements based on other types of specimens, as defined by WHO, and as found in this study

Diagnosis	Venous plasma	WHO		This study		
		Venous blood	Capillary blood	Venous blood	Capillary blood	Venous serum
Impaired fasting glycaemia	6.1	5.6	5.6	5.0	5.6	6.0
Impaired glucose tolerance	7.8	6.7	7.8	6.5	7.2	7.6
Diabetes – fasting	7.0	6.1	6.1	5.8	6.5	6.8
Diabetes – post-challenge	11.1	10.0	11.1	9.4	10.3	10.7

Methodology

Even though the present experiment has not compared all 11 methods to each other directly (only 34 of the 55 possible pairings have actually been tested, cf. Figure 1), the statistical modelling has made it possible to provide conversion formulae for all 55 pairs. While this, from a purely statistical point of view, would be possible with even relatively scanty data, the design of the present study where virtually all samples have been tested by methods n.plas1 and n.plas2 ensures that all other methods have been compared with these two common standards. Thus the conversion formulae between two methods not directly compared, e.g. s.serum and m.plas, are based on at least two common measurements, namely those by n.plas1 and n.plas2.

Other recent studies have mainly compared self-monitoring instruments with a laboratory standard,^{12,13} i.e. they have considered a number of methods against a gold standard that could be assumed to be roughly error-free. This study had a different goal: to provide conversion formulae between methods, allowing any of the methods to have imprecisions associated with them. The estimates of precision (variance components) given in Table 5 might be higher than expected, because they are to some extent dependent on the design of the study, notably the methods chosen for comparison. This is particularly the case for the estimates of the individual \times method interactions. If, for example, fewer methods based on venous plasma and more methods based on venous whole blood had been included we would most likely have seen an increase in the individual \times method interaction SDs for plasma methods and a decrease for methods based on venous whole blood.

Four specimens were collected from each individual during an OGTT. Thus, glucose concentrations at each time-point varied markedly and independently of the baseline value, which made it possible to include each of them as independent values for the calculations. Replicate measurements of the same blood sample were not done, but this would presumably not have added much to the variance components. The variation between the four time-points in each individual is described by the 'true' values μ_{it} , and hence the interindividual variation that would be observed between two blood samples taken, e.g. 5 min apart is not included in variations used in constructing prediction intervals.

In the model fitting process, we looked into whether a log-transform of the glucose measurements was appropriate. It turned out that on purely statistical grounds, there is no particular reason to prefer one scale to the other. From a practical point of view in this study, the actual conversion formulae (applicable in the range 3–15 mmol/L) would of course look mathematically different, but yield essentially the same predictions in this range when converting from one method to another. The major difference between the two approaches is the shape of the prediction bands, which will vary with glucose level for the log-transformed model but not for the model using the original scale. However the actual differences in these are also quite small in the region where the bulk of the data are, so we do

not have much hard empirical evidence to prefer one method to another.

A large uncertainty was found in the measurements by the methods based on capillary blood; both the residual SD and the individual \times method interaction are two to three times higher than for any of the other methods. Thus, capillary blood seems to have limited value as a diagnostic tool. For the other methods, the residual SDs are in the range 0.1–0.3 mmol/L and the individual \times method interaction SD in the range 0.07–0.2 mmol/L, giving prediction SDs largely in the range 0.25–0.45 mmol/L, corresponding to 95% prediction intervals for conversions of the size ± 0.5 to ± 0.9 mmol/L.

Patient population

The patients selected for this study were all overweight (BMI > 25) and were known to have impaired glucose tolerance. This choice was deliberate, because the aim of the study was to compare the methods of measurement, not to characterize any specific population. The underlying assumption is that the obesity and glucose tolerance of the patients do not *per se* interfere with the results of the methods, the only effect being through the actual glucose levels. The main interest is in separating the impaired fasting glucose (IFG), impaired glucose tolerance (IGT) and DM, so that the focus is on the conversion between methods in the range 5–12 mmol/L. In this range, we believe that it is quite a reasonable assumption.

Consistency with cut-off points from diagnostic criteria

In the WHO consultation recommendations, cut-off points are given for diagnosing DM, IGT and IFG, using either fasting or 2 h postload blood glucose.³ These are given for venous plasma, venous whole blood and for capillary whole blood, and thus represents an official belief in what the conversion should be at specific levels of blood glucose between methods based on these three specimens. As noted, there is a tendency that capillary blood glucose does not decay as fast during a 2 h glucose tolerance test as that in venous blood,¹⁴ which in this study is likely to inflate the variance associated with measurements based on capillary blood. However, we found it impractical to provide conversion formulae between specimens that were specific for different clinical situations. This tendency is reflected in the WHO criteria, as can be seen from Figure 2 and Table 7, where these points are given for the methods based on three of the substances mentioned in the WHO recommendations. It is seen that the points for high values of blood glucose are above the line and those for low values below the estimated conversion line.

There is no such tendency for the points linking the plasma and venous whole blood, where the WHO cut-off points are closer to the identity line than the conversion line we estimated, indicating a closer agreement (i.e. smaller bias) than we actually observed in this study. Using the WHO cut-off points would thus mean that it

was more likely to diagnose a person as diabetic if diagnosis were made on the basis of plasma measurements than if based on venous blood.

Fishman *et al.*¹² used a conversion from plasma to venous blood measurements, which was a multiplication by 0.89 (depending somewhat on the haematocrit value). This is in close agreement with our finding (cf. Table 4, second entry in first column). However, we found an intercept of about -0.5 mmol/L, corresponding to the difference between the estimated relation and the relation implied by the various WHO cut-off points. The cut-off points from the recommendations of the WHO consultation thus seem not to be consistent with our findings of conversion factors between methods based on plasma and venous blood respectively. How large the misclassification problem is in practice will depend on the population where a given test is applied. This cannot be inferred from this study where the study subjects were chosen to give maximal power in method comparisons and not to be representative of any particular population.

Nevertheless, it would be prudent to re-evaluate the issue of cut-off points based on specimens other than plasma. Although it has been proposed a long time ago that glucose determination should be done on plasma rather than whole blood,^{15,16} other types of specimens are used. This is an important cause of confusion for both health professionals and patients that until now has not received sufficient attention. Despite the practical problems associated with plasma measurements, the simplest standardization would be to define DM and glucose intolerance only in terms of measurements on venous plasma. If, for understandable practical and logistic reasons WHO wants to keep the other options, then the need for revising the conversion factors should be considered.

ACKNOWLEDGEMENTS

The skilful technical assistance in the blood sampling of Pirjo Härkönen, Anja Ilmanen, Kaija Kettunen, Paula Nyholm and Arja Putilais is gratefully acknowledged. We are also indebted to the researchers in the Finnish Diabetes Prevention Study Group: Sirkka Aunola, Zygimantas Cepaitis, Johan Eriksson, Martti Hakumäki, Helena Hämäläinen, Pirjo Ilanne-Parikka, Sirkka Keinänen-Kiukaanniemi, Mauri Laakso, Pirjo Lehto, Anne Louheranta, Marjo Mannelin, Vladislav Moltchanov, Merja Rastas, Virpi Salminen, Matti Uusitupa and Timo Valle.

The Finnish Diabetes Prevention Study is partly supported by grants from the Academy of Finland (40758, 46558), Finnish Diabetes Research Foundation, Novo Nordisk Foundation, Juho Vainio Foundation and Yrjö Johansson Foundation.

REFERENCES

- 1 Neely RDG, Kiwanuka JB, Hadden DR. Influence of sample type on the interpretation of the oral glucose tolerance test for gestational diabetes mellitus. *Diabet Med* 1991;8:129-34
- 2 Farrer M, Albers CJ, Neil HA, *et al.* Assessing the impact of blood sample type on the estimated prevalence of impaired glucose tolerance and diabetes mellitus in epidemiological surveys. *Diabet Med* 1995;12:325-9
- 3 WHO. *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus.* WHO: Department of Noncommunicable Disease Surveillance, 1999 (who/ncd/ncs/99.2 edition)
- 4 Eriksson J, Lindström J, Valle T, *et al.* Prevention of type II diabetes in subjects with impaired glucose tolerance: the Diabetes Prevention Study (DPS) in Finland. Study design and 1-year interim report on the feasibility of the lifestyle intervention programme. *Diabetologia* 1999;42:793-801
- 5 Tuomilehto J, Lindström J, Eriksson JG, *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *NEJM* 2001;344:1343-50
- 6 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10
- 7 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Meth Med Res* 1999;8:136-60
- 8 Carstensen B. Comparing and predicting between several methods of measurement. *Biostatistics* 2004;5:399-413
- 9 Burrin JM, Alberti KG. What is blood glucose: can it be measured? *Diabet Med* 1990;7:199-206
- 10 Mager M, Farese G. What is 'true' blood glucose? A comparison of three procedures. *Am J Clin Pathol* 1965;35:104-8
- 11 Sidebottom RA, Williams PR, Kamarek KS. Glucose determination in plasma and serum: potential error related to increased haematocrit. *Clin Chem* 1982;28:190-2
- 12 Frishman D, Arditio DM, Graham SS. Performance of glucose monitors. *Lab Med* 1992;23:179-84
- 13 Poirier JY, le Prieur N, Campion L, *et al.* Clinical and statistical evaluation of self-monitoring blood glucose meters. *Diabetes Care* 1998;21:1919-24
- 14 Whichelow MJ, Wigglesworth A, Cox BD, *et al.* Critical analysis of blood sugar measurements in diabetes detection and diagnosis. *Diabetes* 1967;16:219-26
- 15 Wu H. Separate analyses of the corpuscles and the plasma. *J Biol Chem* 1922;51:21-31
- 16 Folin O. Unlabelled blood as a basis for blood analysis. *J Biol Chem* 1930;86:173-8

(Accepted 4 July 2007)