Another approach to estimating the reliability of glycaemic index

Sheila M. Williams[1]*, Bernard J. Venn[2], Tracy Perry[2], Rachel Brown[2], Alison Wallace[3], Jim I. Mann[2] and Tim J. Green[2]

[1]*Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago, PO Box 913, Dunedin, New Zealand*
[2]*Department of Human Nutrition, University of Otago, Dunedin, New Zealand*
[3]*The New Zealand Institute for Crop and Food Research Limited, Lincoln, New Zealand*

The usefulness of the glycaemic index (GI) of a food for practical advice for individuals with diabetes or the general population depends on its reliability, as estimated by intra-class coefficient (ICC), a measure having values between 0 and 1, with values closer to 1 indicating better reliability. We aimed to estimate the ICC of the postprandial blood glucose response to glucose and white bread, instant mashed potato and chickpeas using the incremental area under the curve (iAUC) and the GI of these foods. The iAUC values were determined in twenty healthy individuals on three and four occasions for white bread and glucose, respectively, and for potato and chickpeas on a single occasion. The ICC of the iAUC for white bread and glucose were 0·50 (95 % CI 0·27, 0·73) and 0·49 (95 % CI 0·22, 0·75), respectively. The mean GI of white bread was 81 (95 % CI 74, 90) with a reliability of 0·27 indicating substantial within-person variability. The GI of mashed potato and chickpeas were 87 (95 % CI 76, 101) and 28 (95 % CI 22, 37) respectively with ICC of 0·02 and 0·40. The ICC of the iAUC were moderate and those of the GI fair or poor, indicating the heterogeneous nature of individuals' responses. The unpredictability of individual responses even if they are the result of day-to-day variation places limitations on the clinical usefulness of GI. If the very different GI of potato and chickpeas are estimates of an individual's every-day response to different foods, then the GI of foods may provide an indication of the GI of a long-term diet.

Glycaemic index: Reliability: Methodology: Human studies

Glycaemic index (GI) is the ratio of postprandial glycaemic response following ingestion of a test and a reference food containing equivalent amounts of available carbohydrate, expressed as a percentage[1]. Originally the GI concept was developed to assist individuals with diabetes in choosing carbohydrate-rich foods[2], as the lower postprandial glycaemic response following low-GI foods was considered to improve overall glycaemic control. Low-GI diets are now often recommended in the general population based on the assumption that they promote a healthy body weight and reduce the risk of several chronic diseases[3]. Glycaemic load (GL), which is calculated as the product of the amount of carbohydrate and GI, has been suggested as a potentially more useful concept than GI since it provides an indication of glycaemic response to a typically consumed quantity of carbohydrate. Positive associations between risk of diabetes and CHD and dietary GI and/or GL have been found in some, but not all, epidemiological data[4]. Thus, despite scientific and public interest in GI and GL, their usefulness in disease prevention and diabetes management remains controversial, although some benefit in blood glucose control for individuals with diabetes is indicated[5].

A criticism of GI is that its measurement lacks accuracy and precision[4]. The GI of foods are usually estimated in a small number of subjects (*n* 6–12) and presented with standard errors typically ranging from 3 to 15, indicating that the CI for the GI of some foods are large[6]. For example, if the sample estimate of GI of a particular food were 70 with a standard error of 10, the population estimate of GI would have a 95 % chance of being between 50 and 90, a very imprecise estimate. The estimate of 70 would be regarded as accurate or valid if it differed from the 'true' or the underlying 'error-free' value of the parameter of interest by an acceptably small amount, and precise if the 95 % CI were narrow. Such estimates are essential if GI is to be used in examining diet and disease relationships[7] or as a strategy in diabetes prevention and management[3,8,9].

If an individual's response to a given food is measured repeatedly, in conditions as uniform as possible, it varies because of random error or because of the inherent lability of the measure. An indication of the consistency or repeatability of these responses in a sample of participants can be obtained using the intra-class coefficient (ICC) of reliability (ICC or reliability for short) described in the Appendix[10]. The reliability of a

measure depends on both the between- and within-person variability. It can have values between 0 and 1. Better reliability, a value closer to 1, is obtained when the within-person variability is small when compared with the between-person variability. A consistent response to a particular food, or minimal within-person variability, is required for good reliability or an ICC of an acceptable level. This is important if GI is to be used as a strategy in the management of diabetes.

If GI is regarded as simply a property of food it seems advantageous to use an individual's response to a reference food to adjust the glycaemic response to a particular food in the interest of reducing the between-person variation. Estimates of the proportion of total variance attributable to the within- and between-person variation show that adjusting for a reference food in this way reduces the proportion of variance attributable to between-person variation at the expense of the within-person variance[11]. The within-person variance of the GI is an indication of the extent to which estimates of GI depend on the participants' day-to-day responses and it or the CV has been used as a measure of repeatability. The CV is a measure of relative variability as it depends on both the mean and the scale of measurement, making comparisons of the CV for the area under the curve (AUC) of a food and its GI, for instance, difficult[12].

The purpose of the present study was to estimate the reliability of the components of GI (the incremental AUC (iAUC)), and of GI itself, by determining the within- and between-person variation in glycaemic response. The reliability is presented as the ICC which has well-recognised criteria for qualitative interpretation[10,13,14].

Subjects and methods

Subjects

Twenty individuals, eleven women and nine men, with a normal response to a 2 h oral glucose tolerance test (OGTT) participated in the study. The age and BMI of the group were 23·3 (SD 3·5) years and 23·4 (SD 3·3) kg/m^2 respectively. The Human Ethics Committee of the University of Otago approved the study and all participants gave informed consent.

Study design

This was part of a larger study carried out over a 4-month period in which participants completed a series of thirty-three tests following a predetermined balanced randomisation of all products whose sequence differed for each individual[15]. The iAUC was calculated for beverages containing 0, 12·5, 25, 50 and 75 g anhydrous glucose, with tests carried out in triplicate except for the 50 g glucose test, which was repeated on four occasions. Five foods were tested at three serving sizes. These included white bread providing 50 g available carbohydrate tested on three occasions, potato providing 50 g available carbohydrate and chickpeas providing 25 g available carbohydrate studied on a single occasion. The iAUC for the repeated tests for the 50 g glucose beverage, a 25 g glucose beverage and white bread providing 50 g available carbohydrate as well as the response to potato, and chickpeas were used to examine the reliability of GI. The amount of available carbohydrate in the test foods was determined by

measuring the total starch and total sugars content of the food. Total starch was determined using Association of Official Analytical Chemists method 996.11[16]. Total sugars were determined as the sum of sucrose, lactose, maltose, glucose and fructose, each measured by a GLC method.

Participants were instructed to consume an evening meal containing a large proportion of carbohydrate-rich food the night before each test. It was suggested that the meal be based on rice, bread or potato. Participants were asked not to drink alcohol or consume food or beverages (other than water) after 22.00 hours the night before the tests. On the mornings of the tests the participants were asked to refrain from physical exercise and to report to the clinic in a fasting state. Capillary blood samples were taken using a lancet. A drop of blood was collected into a HemoCue® cuvette and blood glucose concentration measured using a Hemocue® Glucose 201 Analyzer (Aktiebolaget Leo, Helsingborg, Sweden). Each morning the instrument was checked using its own internal system to confirm that it was functioning correctly. The mean of two fasting blood glucose concentrations determined 5 min apart was used as a baseline measure. The time of the second fasting sample was recorded as the start time of the test. White bread or glucose were consumed at an even pace over a period of 15 min and capillary blood samples were taken at 15, 30, 45, 60, 90 and 120 min after the start time. If blood glucose concentrations were >0·2 mmol/l above baseline at 2 h, further blood samples were taken at 150 and 180 min. Participants were asked to remain seated for the duration of the tests. The iAUC was calculated using the method described in FAO/WHO Food and Nutrition paper 66[9].

Before the study, we carried out a validation of the Hemocue® Glucose 201 Analyzer against our standard method. For the standard method, capillary fingerprick blood was collected in micro-centrifuge tubes containing anticoagulant and held on ice until centrifuged at 3000 rpm at 4°C. Blood plasma was pipetted off the packed erythrocytes and stored at −18°C until analysis. The plasma glucose concentration was analysed using an enzymic UV test with hexokinase (Glucose HK Unikit III; Roche Diagnostica, Basle, Switzerland). All glucose analyses were performed on a Cobas II Fara autoanalyser (Roche Diagnostica). We measured iAUC for glucose (× 3) and two test foods (yoghurt and bread) in twelve test subjects using the Hemocue® analyser and our standard method. Mean GI using the Hemocue method and our standard method for yoghurt were 37 (SE 5) and 37 (SE 4) %, respectively. For bread the corresponding means were 57 (SE 3) and 58 (SE 5) %.

Statistics

The means and standard deviations for all the foods and naive within-subject CV were calculated as the mean of each individual's CV obtained from the standard deviation/mean where there were replicates[6,17]. Because the variance increased as the mean increased the data were log-transformed before analysis. A single-factor repeated-measures ANOVA was used to estimate the between- and within-person standard deviations and the ICC of reliability of the log-transformed iAUC for glucose and white bread and the reliability of the GI of each food. A random coefficient model, which accounted for the correlation between the estimates for each individual using the log-transformed values, was used to

estimate the GI and its 95 % CI for each food and compare the GI of the different foods. Person was used as the random effect. As the values for iAUC were log-transformed the GI of each food was estimated as the difference between the average(log(iAUC test food)) and the average(log(iAUC glucose)). This estimate and its CI were back-transformed to obtain the GI of the test food and its CI.

The reliability of a single observation is defined as $ICC = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$, where σ_B^2 is the between-person variance and σ_W^2 is the within-person variance. The reliability of the mean of m independent replicates is given by: $R_m = mR/(1 - (m - 1)R)$[10].

Any observation made on an individual is unreliable in the sense that if the observation were repeated under similar conditions it would differ to some extent from the first. If many hypothetical replicate measurements were made under as close to uniform conditions as possible the mean of these measurement would represent an individual's 'true score' or underlying 'error-free score' for a particular characteristic[10]. The ICC of an observation is estimated in the context of a study where each participant is measured several times[10]. The sample estimate of the mean GI and its reliability can be used to estimate the 'true underlying GI' or error-free estimate of each individual's GI as well as its 95 % CI[18,19] (see Appendix).

The data were analysed with Stata Release 9 (StataCorp LP, College Station, TX, USA)[20].

Results

The means, standard deviations, geometric means and ranges of the iAUC for glucose and the three foods tested are shown in Table 1. All participants consumed 118 g white bread (50 g available carbohydrate) on three occasions, and a beverage containing 50 g glucose on four occasions. The data, ordered according to participants' mean glucose iAUC, are illustrated in Fig. 1. The postprandial glycaemic responses in some individuals are reasonably consistent, while in others day-to-day responses are markedly different. For example, the minimum and maximum iAUC for participant no. 8 were comparatively consistent, 83–121 for three bread tests, and 151–200 mmol/l × min for four glucose tests. For participant no. 14, the ranges for bread and glucose were wider, 83–218 and 167–287 mmol/l × min, respectively. There was seldom a consistent separation in iAUC between white bread and glucose, with considerable overlap occurring in some cases, for example participants no. 3 and no. 5.

The data in Fig. 1 show that the variability of the iAUC values for both glucose and white bread increase as the mean increases. The data were log-transformed so that the variance did not increase with the mean. The mean and standard deviations (on the log scale) are shown in Table 2. The between- and within-person standard deviations for glucose, also shown in Table 2, were 0·27 and 0·28, indicating that the reliability of a single observation of the iAUC is 0·50 (95 % CI 0·27, 0·73), and 0·66, 0·75 and 0·80 for the mean of two, three and four replicates, respectively. The between- and within-person standard deviations for white bread were 0·29 and 0·30. The reliability of the iAUC values for a single measure of white bread was 0·49 (95 % CI 0·22, 0·75), and 0·65 and 0·74 for the mean of two and three replicates, respectively. The estimate of the GI of white bread was 81 (95 % CI 74, 90). The reliability or ICC of the GI based on four replicates of glucose and three of white bread was 0·27, from which between- and within-standard deviations of 0·13 and 0·22 on the log scale, respectively, can be inferred.

The GI for white bread and its 95 % CI (width 16 units) and the GI values for each individual are shown in Fig. 1. Interestingly, the participant with the highest GI had one of the lowest and most consistent responses to both glucose and white bread. If GI is regarded as an estimate of an individual's response to white bread, with random day-to-day fluctuations, future estimates would be expected to be closer to the mean. The width of the 95 % CI for an individual's score, centred on the 'true' GI (the GI measured without error), ranged from 32 to 44. These differ from individual to individual because the variance increases as the mean increases.

The data for each individual for both potato and the 50 g glucose beverage are presented in Fig. 2 (A). The individual values for the GI along with the sample mean and 95 % CI are shown Fig. 2 (B). The GI for potato was 87 (95 % CI 76, 101). The reliability, assuming that the reliability of the iAUC of potato was similar to that obtained for glucose, was 0·02. This means that the wide CI for the GI of potato can be attributed almost entirely to within-person variability. The 'true score' for individuals depends on the sample mean. The ratio of the GI of potato to white bread was 1·07 (95 % CI 0·88, 1·31) and not statistically different ($P=0·47$) from 1.

The iAUC values for glucose (25 g) and chickpeas are shown in Fig. 3. There is much less overlap between the values in this case. The reliability for the log-transformed values for glucose was 0·47 (95 % CI 0·21, 0·74), with between- and within-person standard deviations of 0·28 and 0·27 respectively. The GI for chickpeas was 28 (95 % CI 22, 37).

Table 1. Incremental area under the curve (mmol/l × min) for glucose and foods

(Mean values and standard deviations for twenty subjects)

	Replicates	Mean	SD*	CV (%)†	Geometric mean	Range
Glucose (50 g)	4	211	69·2	25·3	195	114, 332
White bread (50 g‡)	3	173	63·2	25·9	159	82, 344
Potato (50 g‡)	1	187	85·3	NE	171	63, 376
Glucose (25 g)	3	137	41·9	23·7	127	76, 220
Chickpeas (25 g‡)	1	49	40·0	NE	30	0, 156

NE, not estimatable.

* The standard deviation is based on the distribution across individuals after averaging the within-person replicates.
† Calculated as the mean of each individual's CV obtained from the standard deviation/mean where there were replicates.
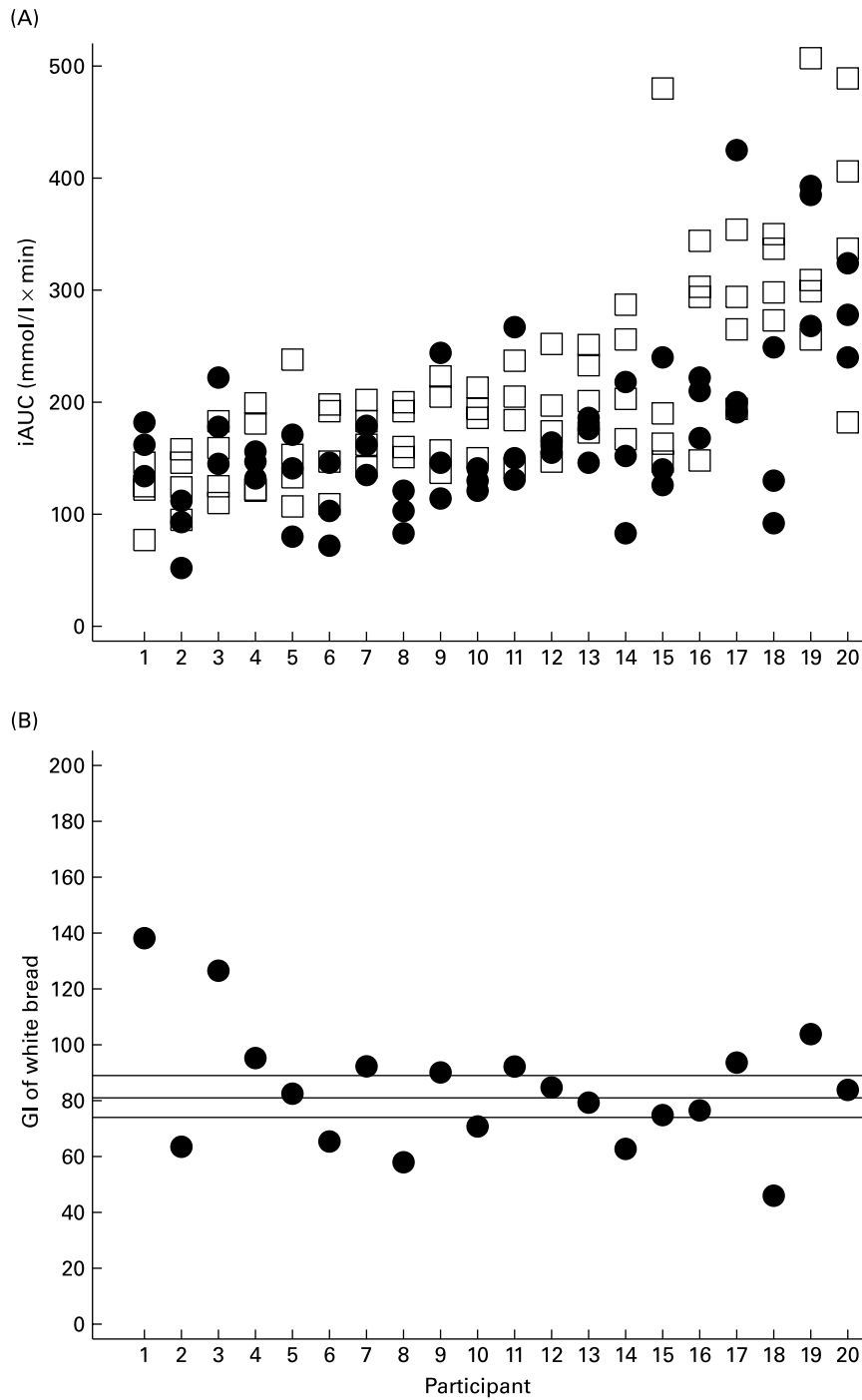‡ Available carbohydrate.

(A)

(B)

Fig. 1. (A) The incremental blood glucose area under the curve (iAUC; mmol/l × min) estimates for three replicates of 118 g white bread (●) and four replicates of 50 g glucose beverages (□) measured in twenty participants. (B) The glycaemic index (GI) of white bread for each individual (●) and the mean GI and 95 % CI for the sample (——). The data have been ordered by mean of values for glucose.

The reliability or ICC of the GI of chickpeas was 0·40, assuming that the reliability of the iAUC of chickpeas was the same as that for 25 g glucose. The mean and its 95 % CI (width 15) are shown in Fig. 3 along with the estimate of GI for each individual. Although future values of the GI for chickpeas for an individual would be pulled towards the population mean, the width of the 95 % CI for an individual's score would be between 16 and 97. Thus, a wide range of plausible values for individuals'

responses is possible. The ratio of the GI of chickpeas to white bread was 0·34 (95 % CI 0·27, 0·44), which was statistically significant ($P < 0.001$).

Discussion

The measures of reliability of the GI of foods described in the present report indicate that their repeatability was unacceptably

Table 2. Means, standard deviations and reliability for log-transformed values of the incremental area under the curve for glucose and white bread for twenty subjects

	Glucose (50 g)	White bread
Replicates	4	3
Log-transformed scores		
Mean	5·3	5·1
SD	0·31	0·34
Within-person SD	0·28	0·30
Between-person SD	0·27	0·29
Reliability of a single replicate	0·50	0·49
Reliability of a single replicate 95 % CI	0·27, 0·73	0·22, 0·75
Reliability using all replicates	0·80	0·74
Reliability using all replicates 95 % CI	0·61, 0·91	0·45, 0·89

low. The ICC, which has been recommended for use in human nutrition[12], indicates that the repeatability of the iAUC used in the present study was moderate and those of GI fair to poor[10,14]. Values of reliability in the region of 0·5 indicate that the within-person and between-person variances are similar in magnitude; fair or poor reliability, as is the case for the GI of potato, is the result of the within-person variability being much larger than the between-person variability. Although the use of a reference food reduced the between-person variability for GI the within-person variability was still considerable. The magnitude of the overall error, accumulated from the large errors associated with the iAUC, is indicated by the wide CI for the GI means. Modest improvements in the reliability of glucose and white bread could be obtained by increasing the number of replicates, though as the law of diminishing returns applies most gain in reliability comes with the first two or three replicates, with little extra being realised from adding further replicates.

The iAUC in the present study were based on glucose measurements taken at the recommended time points. Increasing the frequency of the observations would capture the glucose profile more accurately and could increase its reliability. More frequent observations would also mean that other ways of summarising the glucose profile such as the maximum glucose concentration could be considered[21]. However, more frequent observations would add to the cost and perhaps make recruiting and retaining volunteers more difficult. The reliability of GI could be affected by the time interval between repeat tests. If the GI of a food is be used for practical advice in real life, however, it is important that its reliability does not depend on the time interval between the tests. If population estimates are the main concern, problems created by poor reliability could be overcome by increasing the sample size.

The data for the twenty participants shown in Fig. 1 (A) illustrate both the considerable overlap and the wide variation for the values of the iAUC for glucose and white bread. The wide CI indicating that plausible values of the population mean for the GI of white bread were between 74 and 90 were a consequence of this. The CI for the GI of potato and chickpeas were also wide, making the ranking of foods according to their GI, a cornerstone for delivering advice to the population at large, extremely difficult. The CI for individuals' values for the GI of white bread were also very broad,

indicating a very unstable response in some participants. If the variability can be attributed to day-to-day random fluctuations, a larger number of replicates would minimise the effect of the within-person variance on the overall variance, and provide more precise CI for the estimate of the GI of a particular food. In real life an individual's response to a food would average out over many eating occasions, so the estimates of the GI of foods such as potato and chickpeas used in the present study may be an indicator of the GI of a long-term diet. If, on the other hand, the different iAUC values for glucose and the test food mark genuine differences in physiological response to carbohydrate on different days, the validity of GI and its usefulness in clinical practice must be questioned. As the present study was based on a relatively homogeneous group of young individuals it is possible that the variability could be even larger if the participants were older, of different ethnic origin or diabetic.

Food is often classified as low, medium or high GI for convenience[22]. The error inherent with GI measurement based on small samples means that misclassification is likely to occur. In an interlaboratory study, rice would have been classified as low by one laboratory, medium by three, and high GI by three laboratories[6]. Assigning food a category of GI using small samples implies a level of precision unwarranted by the experimental procedure.

The reliability of an OGTT in 111 participants without diabetes in a longitudinal study of an elderly population in the Netherlands was 0·38 for the iAUC, a value comparable with that obtained in the present study[23]. That study also reported a large within-person standard deviation of 87·5 mmol/l × min for the iAUC for the non-diabetic participants. Although the mean value of an OGTT could be different in an older population the reliability, which is a property of the test, should be comparable. Several studies have shown that the reproducibility of OGTT tests is only moderate[24,25] and an article arguing that the OGTT is superfluous suggests that its most serious drawback from the clinical point of view is its lack of reproducibility[26]. If the iAUC in studies of GI are analogous to the OGTT tests, we should not be surprised to find that the reliability is unsatisfactory. The reliabilities reported in the present study were similar to those that can be obtained from data presented in another report[27] and other authors have alluded to the problems associated with imprecise tests such as glucose tolerance tests[28].

The mean and standard deviation for AUC of glucose obtained in the present study were similar in magnitude with those of an interlaboratory study and a more recent Finnish study[6,29]. Individuals' GI for white rice ranged from 60 to 131 in a Japanese study using ten participants, a range in estimates of 71 GI units[30], comparable in magnitude with those in the present study. The Japanese study excluded the largest value because it was more than two standard deviations above the mean although there was no evidence to suggest that the responses to glucose and white rice were not genuine. We did not exclude the participant with the highest GI value from our calculations for the GI of bread because replicate tests of glucose and white bread showed very good consistency in iAUC (Fig. 1 (A); participant no. 1), indicating that the GI calculated for this participant was probably a valid representation of their physiological response.
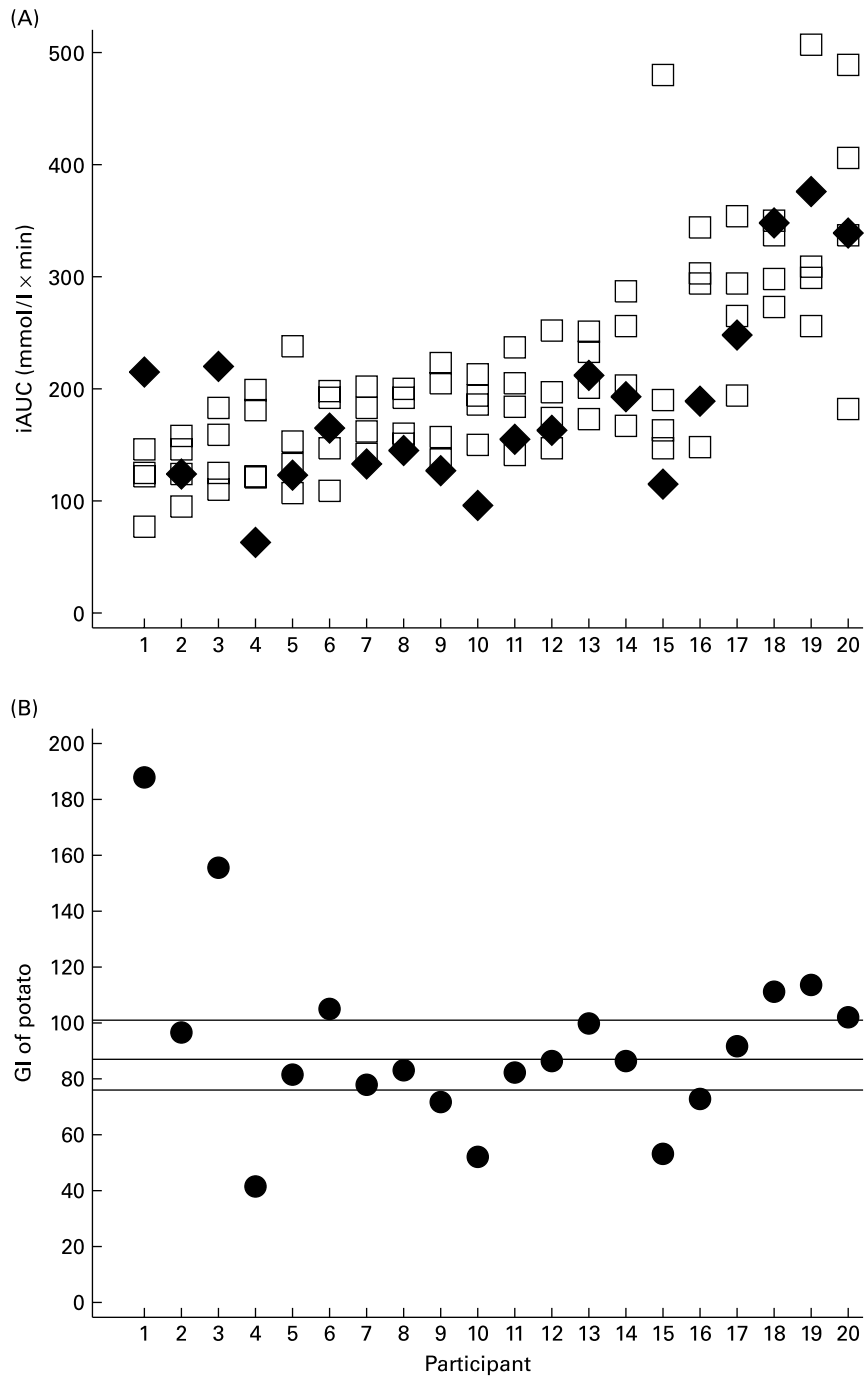
Fig. 2. (A) The incremental blood glucose area under the curve (iAUC; mmol/l × min) estimates for potato (◆) and four replicates of 50 g glucose beverages (□) measured in twenty participants. (B) The glycaemic index (GI) of potato for each individual (●) and the mean GI and 95 % CI for the sample (——). The participants are in the same order as Fig. 1.

The present study included twenty participants rather than the recommended six to twelve participants. It also used log-transformed values to overcome the strong relationship between the mean and the standard deviation, a practice recommended for AUC[21,31]. This led to a more normal distribution of the data, a requirement of statistical tests such as ANOVA and t tests[32]. Reliability coefficients as well as estimates of the within-person CV have also been presented.

It has been suggested that the CV, the ratio of the standard deviation to the mean, provides a useful statistic for comparing the precision of different variables[33]. It is also understood that large CV, as is the case in this and other studies of iAUC, imply imprecise measurement methods. Stronger arguments are made for the use of the ICC, which relates the between-person variance to the total variance comparing like with like as it were[10,12]. The reliability of GI is poor partly
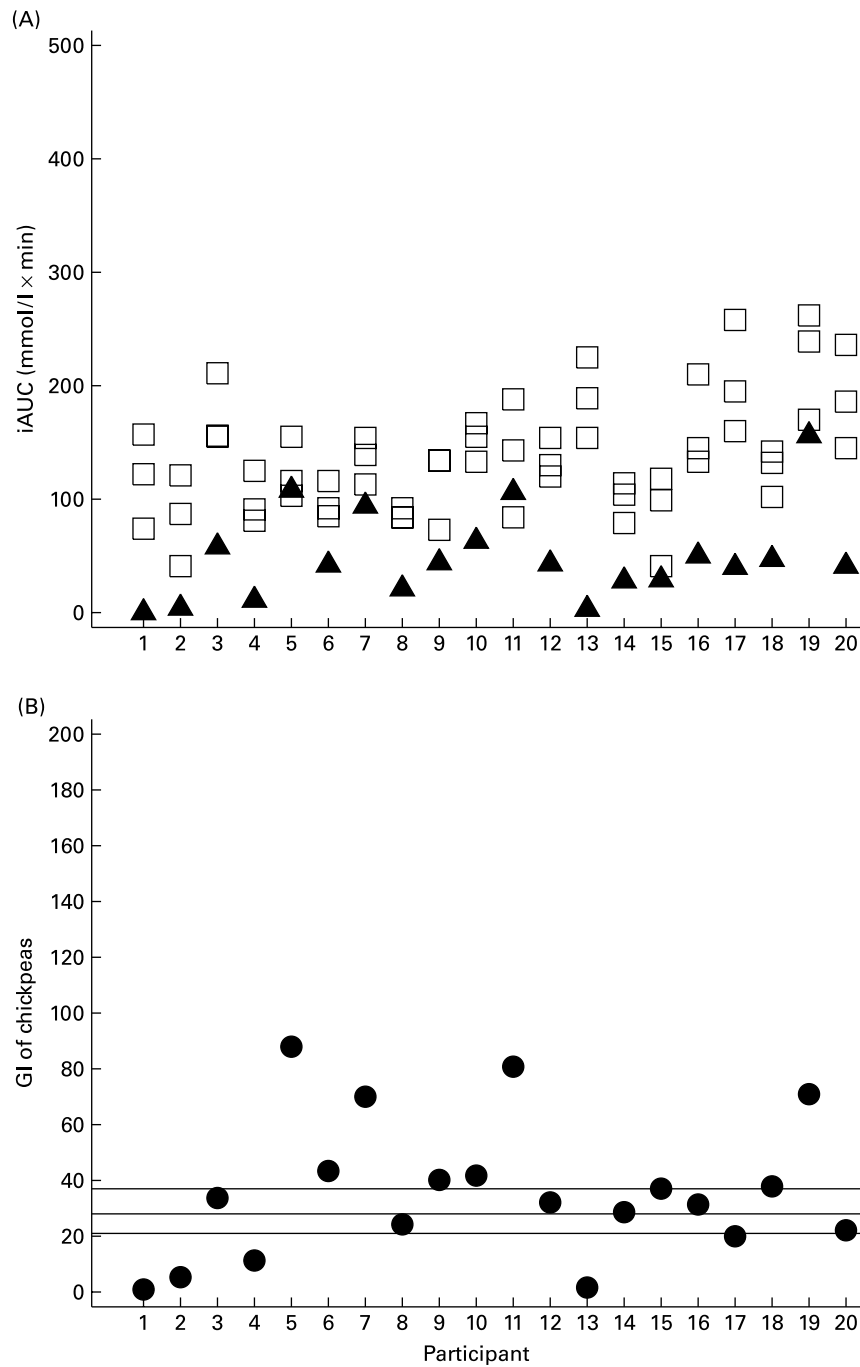
Fig. 3. (A) The incremental blood glucose area under the curve (iAUC; mmol/l × min) estimates for replicates of chickpeas (▲) and three replicates of 25 g glucose beverages (□) measured in twenty individuals. (B) The glycaemic index (GI) of chickpeas for each individual (●) and the mean GI and 95 % CI for the sample (——). The participants are in the same order as Fig. 1.

because the reliability of the iAUC for glucose and foods is only moderate, although this can be improved by replication. Furthermore, individuals' responses to glucose and food are correlated, all be it modestly, which makes the total variance less than it would be if the measures were independent.

In many studies poor reliability is overcome by using very large samples, as is the case in many epidemiological studies, which have examined the relationship between GI or GL and the onset of diabetes or CHD. Larger samples with more

replicates would provide more precise and possibly more accurate estimates of the GI of foods, making ranking and classifying foods more dependable. Nonetheless, as samples of more than thirty subjects using three replicates for both foods are required to provide CI within 10 % of the GI esti-mate, it is unrealistic to think that foods with GI differing by 15 or 20 points, derived from samples of six to twelve sub-jects are necessarily different. The results of the present study do show that the GI of chickpeas is one-third that of white

bread, and that this difference is statistically significant, so GI estimated from small samples may be a useful way of signalling major differences in foods.

One of the practical problems associated with small samples and poor reliability is varying estimates of GI even for the same food. In an interlaboratory study in which the same foods were supplied to GI-testing laboratories each using eight to twelve subjects and three replicates for glucose with no repeat testing of the food, mean GI values ranged from 86 to 99 (instant potato), 55 to 85 (rice), 39 to 70 (spaghetti) and 25 to 46 (barley)[6]. Similarly, in the international tables of GI and GL, the GI of Russet baked potatoes (item 603) are reported to range from 56 to 111[34]. The reasons for such large differences in GI for the same food are not understood, but improving reliability by increasing the number of replicates and sample size should help to reduce differences.

The GI and GL concepts have been used in several contexts. Prospective epidemiological studies have examined the relationship between the GL of diet and subsequent chronic disease. Clinical trials demonstrating the potential of low-GI foods to favourably influence lipoprotein profiles and glycaemic control in individuals with diabetes form the basis of recommendations to encourage the use of low-GI foods. The large number of individuals in prospective studies may justify conclusions relating to the protective role of low-GI foods, despite the lack of reliability demonstrated here. On the other hand the low reliability suggests that individual responses are likely to be unpredictable when using GI to guide food choices in the clinical context. The use of GI in dietary prescriptions may be further limited by the variable nature of foods.

References

1. Wolever TM, Jenkins DJ, Jenkins AL & Josse RG (1991) The glycemic index: methodology and clinical implications. *Am J Clin Nutr* **54**, 846–854.
2. Jenkins DJ, Wolever TM, Taylor RH, Barker H, Fielden H, Baldwin JM, Bowling AC, Newman HC, Jenkins AL & Goff DV (1981) Glycemic index of foods: a physiological basis for carbohydrate exchange. *Am J Clin Nutr* **34**, 362–366.
3. Ludwig DS (2003) Dietary glycemic index and the regulation of body weight. *Lipids* **38**, 117–121.
4. Pi-Sunyer FX (2002) Glycemic index and disease. *Am J Clin Nutr* **76**, 290S–298S.
5. Sheard NF, Clark NG, Brand-Miller JC, Franz MJ, Pi-Sunyer FX, Mayer-Davis E, Kulkarni K & Geil P (2004) Dietary carbohydrate (amount and type) in the prevention and management of diabetes: a statement by the American Diabetes Association. *Diabetes Care* **27**, 2266–2271.
6. Wolever TM, Vorster HH, Björck I, *et al.* (2003) Determination of the glycaemic index of foods: interlaboratory study. *Eur J Clin Nutr* **57**, 475–482.
7. Salmeron J, Manson JE, Stampfer MJ, Colditz GA, Wing AL & Willett WC (1997) Dietary fiber, glycemic load, and risk of non-insulin-dependent diabetes mellitus in women. *JAMA* **277**, 472–477.
8. Kalergis M, Pytka E, Yale JF, Mayo N & Strychar I (2006) Canadian dietitians' use and perceptions of glycemic index in diabetes management. *Can J Diet Pract Res* **67**, 21–27.
9. Anonymous (1998) Carbohydrates in human nutrition. Report of a Joint FAO/WHO Expert Consultation. *FAO Food Nutr Pap* **66**, 1–140.
10. Fleiss JL (1986) *The Design and Analysis of Clinical Experiments*. New York: John Wiley and Sons.
11. Wolever T (2006) *The Glycaemic Index: a Physiological Classification of Dietary Carbohydrate*. Wallingford, UK: CABI.
12. Allison DB (1993) Limitations of coefficient of variation as index of measurement reliability. *Nutrition* **9**, 559–561.
13. Bartko JJ & Carpenter WT Jr (1976) On the methods and theory of reliability. *J Nerv Ment Dis* **163**, 307–317.
14. Dunn G (1992) Design and analysis of reliability studies. *Stat Meth Med Res* **1**, 123–157.
15. Venn BJ, Wallace AJ, Monro JA, Perry T, Brown R, Frampton C & Green TJ (2006) The glycemic load estimated from the glycemic index does not differ greatly from that measured using a standard curve in healthy volunteers. *J Nutr* **136**, 1377–1381.
16. Association of Official Analytical Chemists (1998) *Method 996.11: Starch (Total) in Cereal Products, Amyloglucosidase–α-Amylase Method (Total Starch Assay)*. Gaithersburg, MD: AOAC International.
17. Quan H & Shih WJ (2006) Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics* **52**, 1195–1203.
18. Cook NR (1996) Estimating predictive values for blood pressure measurements from multivariate regression models with covariates. *Stat Med* **15**, 2013–2028.
19. Irwig L, Glasziou P, Wilson A & Macaskill P (1991) Estimating an individual's true cholesterol level and response to intervention. *JAMA* **266**, 1678–1685.
20. StataCorp (2005) *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP.
21. Matthews JN, Altman DG, Campbell MJ & Royston P (1990) Analysis of serial measurements in medical research. *BMJ* **300**, 230–235.
22. Brand-Miller J, Wolever TM, Foster-Powell K & Colaguiri S (2003) *The New Glucose Revolution*. New York: Marlowe & Company.
23. Feskens EJ, Bowles CH & Kromhout D (1991) Intra- and inter-individual variability of glucose tolerance in an elderly population. *J Clin Epid* **44**, 947–953.
24. McDonald GW, Fisher GF & Burnham C (1965) Reproducibility of the oral glucose tolerance test. *Diabetes* **14**, 473–480.
25. Mooy JM, Grootenhuis PA, de Vries H, Kostense PJ, Popp-Snijders C, Bouter LM & Heine RJ (1996) Intra-individual variation of glucose, specific insulin and proinsulin concentrations measured by two oral glucose tolerance tests in a general Caucasian population: the Hoorn Study. *Diabetologia* **39**, 298–305.
26. Davidson MB (2002) Counterpoint: the oral glucose tolerance test is superfluous. *Diabetes Care* **25**, 1883–1885.
27. Wolever TM, Csima A, Jenkins DJ, Wong GS & Josse RG (1989) The glycemic index: variation between subjects and predictive difference. *J Am Coll Nutr* **8**, 235–247.

28. Levy J, Morris R, Hammersley M & Turner R (1999) Discrimination, adjusted correlation, and equivalence of imprecise tests: application to glucose tolerance. *Am J Physiol* **276**, E365–E375.
29. Hatonen KA, Simila ME, Virtamo JR, Eriksson JG, Hannila M-L, Sinkko HK, Sundvall JE, Mykkanen HM & Valstra LM (2006) Methodologic considerations in the measurement of glycemic index: glycemic response to rye bread, oatmeal porridge, and mashed potato. *Am J Clin Nutr* **84**, 1055–1061.
30. Sugiyama M, Tang AC, Wakaki Y & Koyama W (2003) Glycemic index of single and mixed meal foods among common Japanese foods with white rice as a reference food. *Eur J Clin Nutr* **57**, 743–752.
31. Keene ON (1995) The log transformation is special. *Stat Med* **14**, 811–819.
32. Bland JM & Altman DG (1996) The use of transformation when comparing two means. *BMJ* **312**, 1153.
33. Hulley SB & Cummings SR (1988) *Designing Clinical Research*. Baltimore, MD: Williams & Wilkins.
34. Foster-Powell K, Holt SH & Brand-Miller JC (2002) International table of glycemic index and glycemic load values: 2002. *Am J Clin Nutr* **76**, 5–56.
35. Peter JP, Churchill G Jr & Brown TJ (1993) Caution in the use of difference scores in consumer research. *J Consum Res* **19**, 655–662.

Appendix

To quantify the within- and between-person variance of a measure we considered the classical linear model:

$$X = T + e,$$

where X is the observed value for a particular person, T the error-free score and e the error or the difference between the observed and error-free scores. In a population the error-free score T will vary about some mean with a variance of σ_T^2 (between-person variation). For a particular person the random error e varies about a mean zero (within-person variation). Assuming that the distribution of the error is independent of T, e has variance σ_e^2 regardless of the value of T so the variance of X is:

$$\sigma_X^2 = \sigma_T^2 + \sigma_e^2.$$

The reliability of X is defined as[10]:

$$ICC = \sigma_T^2/(\sigma_T^2 + \sigma_e^2).$$

Thus, the reliability of a measure depends on both between- and within-person error and will be high when σ_e^2 (the within-person error) is small in comparison with σ_T^2 (the between-person error). The within-person variability can be controlled by replication as the variance of the mean of m replicates is:

$$\sigma_X^2 = \sigma_T^2 + \sigma_e^2/m.$$

So the reliability of the average of m replications is[10]:

$$ICC_m = \sigma_T^2/(\sigma_T^2 + \sigma_e^2/m).$$

This can also be expressed as:

$$R_m = mR/(1 - (m - 1)R).$$

The reliability of the difference between two measures, in this case averages of the log-transformed values of the AUC for the test food and glucose is:

$$R = (R_1\sigma_1^2 + R_2\sigma_2^2 - 2r_{12}\sigma_1\sigma_2)/(\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2),$$

where σ_1^2 and σ_2^2 are the variances of the two measures, r_{12} is the correlation between them, and R_1 and R_2 are their reliabilities[35].

The 'true score' or the score, which would be obtained if there were no error of measurement, and its standard error, are estimated as follows[18,19]:

$$\text{Estimated 'true score' } T = \mu + R(X_i - \mu),$$

where X_i is an observation from an individual and μ is the sample mean with variance:

$$Var(T) = R(1 - R)\sigma^2.$$