Measuring the glycemic index of foods: interlaboratory study1–5


ABSTRACT
Background: Many laboratories offer glycemic index (GI) services.

Objective: We assessed the performance of the method used to measure GI.

Design: The GI of cheese-puffs and fruit-leather (centrally provided) was measured in 28 laboratories (n = 311 subjects) by using the FAO/WHO method. The laboratories reported the results of their calculations and sent the raw data for recalculation centrally.

Results: Values for the incremental area under the curve (AUC) reported by 54% of the laboratories differed from central calculations. Because of this and other differences in data analysis, 19% of reported food GI values differed by >5 units from those calculated centrally. GI values in individual subjects were unrelated to age, sex, ethnicity, body mass index, or AUC but were negatively related to within-individual variation (P = 0.033) as expressed by the CV of the AUC for repeated reference food tests (refCV). The between-laboratory GI values (mean ± SD) for cheese-puffs and fruit-leather were 74.3 ± 10.5 and 33.2 ± 7.2, respectively. The mean laboratory GI was related to refCV (P = 0.003) and the type of restrictions on alcohol consumption before the test (P = 0.006, r² = 0.509 for model). The within-laboratory SD of GI was related to refCV (P < 0.001), the glucose analysis method (P = 0.010), whether glucose measures were duplicated (P = 0.008), and restrictions on dinner the night before (P = 0.013, r² = 0.810 for model).

Conclusions: The between-laboratory SD of the GI values is ≈9. Standardized data analysis and low within-subject variation (refCV < 30%) are required for accuracy. The results suggest that common misconceptions exist about which factors do and do not need to be controlled to improve precision. Controlled studies and cost-benefit analyses are needed to optimize GI methodology. The trial was registered at clinicaltrials.gov as NCT00260858. Am J Clin Nutr 2008;87(suppl):247S–57S.

KEY WORDS Clinical trial, humans, dietary carbohydrate, glycemic index, glucose, methodology

INTRODUCTION

The glycemic index (GI) is a measure of the blood glucose-raising ability of the available carbohydrate in foods defined as the incremental area under the glycemic response curve (AUC) elicited by a portion of food containing 50 g available carbohydrate expressed as a percentage of the AUC elicited by 50 g glucose in the same subject. Prospective studies suggest that low-GI diets may reduce the risk of diabetes (1, 2), cardiovascular disease (3, 4), metabolic syndrome (5), chronic inflammation (6), and possibly some types of cancer (7–12). Clinical trials have shown that low-GI diets improve glycemic control in diabetes (13), increase insulin sensitivity (14, 15) and β-cell function (16, 17), reduce food intake (18) and body weight (19–21), influence memory (22, 23), and may reduce serum cholesterol (24). Diabetes associations in the United Kingdom (25), Canada (26), Australia (27), Europe (28), and the United States (29) indicate that GI is a useful tool for differentiating between carbohydrates. For these reasons, labeling of GI on foods has been proposed or is already occurring in Australia, South Africa, Sweden, United Kingdom, and Germany, with several commercial laboratories measuring the GI of foods.

For regulatory purposes, an approved method of measuring the GI of foods is required, and standards must be developed to

1 From the Department of Nutritional Sciences, University of Toronto and Glycemic Index Laboratories, Inc, Toronto, Ontario, Canada (TMSW); the Human Nutrition Unit, Department of Biochemistry, University of Sydney, Sydney, NSW, Australia (JCB-M, FA, and GSD); AMK Research, Inc, Gainsville, FL, USA (JA); the Department of Human Nutrition, Centre for Advanced Food Studies, Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark (AA and BS); the Department of Clinical Nutrition, Gothenburg University, Gothenburg, Sweden (MA and LD); the Centre for Chemistry and Chemical Engineering, Lund University, Lund, Sweden (IB and YG); the Human Nutrition Unit, Department of Public Health, University of Parma, Parma, Italy (FB and FS); the Department of Human Nutrition, University of Otago, Dunedin, New Zealand (RB and TP); the Nutrition and Dietetic Department, Hammersmith Hospital, Imperial College, London, United Kingdom (AB and GF); DISTAM - Nutrition Unit, University of Milan, Milan, Italy (MCC and DE); Biofortis, Nantes, France (MC); the GI Foundation of South Africa, Mpumalanga, South Africa (ED); the Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, United Kingdom (SHampton, KLI, and LMorgan); the Nutrition Unit, Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute (KTL), Helsinki, Finland (KH and LMV); the School of Life Sciences, Oxford Brookes University, Oxford, United


247S
enable assessment of the performance of the laboratories. The effect of some methodologic variables on GI values is known, and a recommended method is available (30, 31). However, the method does not address all common methodologic variations. A previous interlaboratory study suggested that the between-laboratory SD of average GI values of the same food determined in 8–10 subjects was ≈9 (32). However, too few centers were involved (n = 7) to reliably assess the extent of variation in methodology in different laboratories around the world and the effects of such variables on the results obtained. In addition, the extent of variation in data analysis by individual laboratories was not assessed, because all calculations were performed centrally. Thus, the purposes of this study were to assess the magnitude of the variation of the means and SDs of GI values measured by different laboratories around the world and to determine the extent to which sources of methodologic variation may explain the interlaboratory variation of GI means and SD.

### MATERIALS AND METHODS

Each participating laboratory was sent 3 food products for GI determination: an oat biscuit, cheese puffs, and fruit leather (Table 1). As a result of a misunderstanding about the definition of the term available carbohydrate between the investigators and the manufacturer, it was subsequently learned that the portion size of the oat biscuit used contained 40% more available carbohydrate than expected, and so the results for this food are not included here. The foods were chosen because they were ready-to-eat, and preliminary data suggested that one would have a high GI (>70) and the other a low GI (≤55). The protocol indicated that 10 healthy subjects should be studied in each location, but some laboratories included up to 14 subjects. The subjects were males and nonpregnant, nonlactating females aged 18–75 y who did not have diabetes. The protocol used by each location was approved by a human subjects ethics review committee, and each participating subject provided informed consent by signing an approved consent form.

In each location, the subjects were studied on 3–5 separate occasions in the morning after they had fasted overnight. On 2 occasions, they consumed test meals consisting of one of the test foods, and on the other occasions, the test meal consisted of the reference food. On each occasion, after a fasting blood sample was taken, the subjects ate the test meal and had further blood samples drawn at 15, 30, 45, 60, 90, and 120 min after starting to eat. The protocol indicated that capillary blood should be obtained by finger-prick, but one site obtained venous blood from a forearm vein, and one site obtained capillary blood from the earlobe.

The portion size of each test food contained 50 g available carbohydrate (defined as total carbohydrate minus dietary fiber) based on the information on the nutrition information panel. To avoid misunderstanding about the definition of the terms on the food label, all sites were advised about the portion sizes of the test foods to be used. The reference food could be 50 g anhydrous glucose, 55 g dextrose (glucose-monohydrate), or 50 g available carbohydrate from white bread. Each laboratory determined the number of times the reference food was tested by each subject. The protocol indicated that each test meal was to be served with 250 to 500 mL water or tea (50 mL milk was allowed if desired). Each subject could choose the volume and type of drink desired, but the drink chosen was the same for all test meals consumed by that subject. Most sites restricted the drink to only water, and one
allowed coffee as a choice. Test meals were consumed within 10 min. Timing for blood samples started with the first bite of the test meal.

Each laboratory could measure glucose in whole blood, plasma, or serum by any recognized method, as long as the method used was the same for all tests. The incremental area under the glucose response curve (AUC) above the fasting glucose concentration, ignoring the area beneath the fasting concentration, was calculated by using the trapezoid rule. The AUC of each subject after each test food was expressed as a percentage of the mean AUC elicited by the reference food in the same subject. The mean of these values for all the subjects was the food GI. If white bread was used as the reference food, the GI values were multiplied by 0.71 to convert them to the glucose scale (ie, the GI of glucose = 100).

Participating laboratories were asked to send the following information about their methods and results to the central laboratory (University of Toronto): inclusion and exclusion criteria for subjects; the reference food used; the method used to sample blood and measure glucose; whether measures of glucose were repeated; the restrictions (if any) placed on subjects with respect to exercise, alcohol consumption, smoking, length of fasting, and the meal consumed the evening before each test; the drink served with the test meal; the age, sex, ethnicity, height, weight, and body mass index (BMI; in kg/m²) of each subject; medications used by subjects (if any); glucose concentrations at each time point; AUC values calculated locally for each test meal taken by each subject; and the mean and SD of the GI values calculated locally.

The adiposity of white subjects was classified as underweight, ideal weight, overweight, or obese by using BMI cutoffs of <18.5, 18.5–25.0, 25.1–30.0, and >30.0, respectively. Because, for a given BMI, South and East Asians have more body fat than do whites (33), nonwhite subjects were classified as underweight, ideal weight, overweight, or obese by using BMI cutoffs of <18.0, 18.0–23.0, 23.1–26.9, and >26.9.

The mean and SD of the AUC and GI values reported by each laboratory were compared with those calculated at the central laboratory by using the procedure of Bland and Altman (34). The critical value for this procedure is termed limits of agreement, defined as the mean ± 2 × SD of the differences between the results using 2 different methods (in this case, the values reported by each laboratory and the respective values calculated by the central laboratory); thus, the limits of agreement represent the range within which 95% of the differences lie. Limits of agreement for AUC ±2 mmol·min/L and for mean GI ±1.0 were considered to be due to rounding and, therefore, insignificant; larger values were considered to be due to calculation or reporting errors. Because significant differences were found between reported and centrally calculated values for AUC and GI, centrally calculated AUC and GI values were used for further statistical analysis. The mean and CV (CV = 100 × SD/mean) of the AUC values for repeated reference food tests were calculated for each subject, and the results were termed reference AUC and reference CV (refCV), respectively (CV could be calculated only for the 26 laboratories in which the reference food was tested more than once by each subject).

GI values were calculated by expressing each subject’s AUC after the test food as a percentage of the same subject’s mean reference AUC. The mean of the resulting values was the GI of the food. Within each laboratory, individual values greater than the mean plus 2 SDs were considered to be outliers and were excluded from the final results.

Individual AUC and GI values for the 2 test foods were compared by analysis of variance (ANOVA) for a 2-factor experiment with repeated measures on one factor (35) examining the main effects of laboratory and food and the laboratory × food interaction. The influence of laboratory methods on laboratory mean and SD of the GI values was determined by step-wise multiple regression analysis (Lotus 1-2-3 97 Edition; Lotus Development Corp., Cambridge, MA) by using the step-up procedure described by Snedecor and Cochran (36). For regression analysis involving subject variables, dummy values were used for sex (P = 0, M = 1) and ethnicity (white = 0, other = 1). For regression analysis involving methodologic variables, the methods used at each laboratory were placed into 2 to 4 categories, depending on the range of methods used, and dummy values between 0 and 3 were assigned to the categories in order of the unadjusted means within the categories. Information about the use of duplicate blood samples and duplicate measurements of glucose was collapsed into a single category whether any duplication was done or not.

## RESULTS

A total of 314 subjects were involved in the study, of whom 2 dropped out and 1 was excluded for being discovered to have impaired glucose tolerance. Of the 311 who were included, 241 (77%) were white, 26 were Southeast-Asian, 21 were East-Asian, 16 were South-Asian, 3 were African, 3 were mixed African/South-Asian, and 1 was Middle-Eastern (Table 2). The white subjects did not differ significantly from the others in sex or age, but were significantly taller and heavier; however, the difference in BMI was not significant. The glycemic response elicited by the reference food was significantly less in the white subjects than in the others (Table 2), but the reference CV and mean GI values did not differ significantly. Fourteen laboratories provided information about the subjects’ medication use; of the

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>White (n = 241)</th>
<th>Other (n = 70)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males/females (n)</td>
<td>101/140</td>
<td>26/44</td>
<td>NS</td>
</tr>
<tr>
<td>Age (y)</td>
<td>30.0 ± 0.7²</td>
<td>28.7 ± 1.1</td>
<td>NS</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>172 ± 1</td>
<td>163 ± 1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>69.9 ± 0.9</td>
<td>60.3 ± 1.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.4 ± 0.2</td>
<td>22.6 ± 0.4</td>
<td>0.090</td>
</tr>
<tr>
<td>Underweight [n (%)]</td>
<td>4 (2)</td>
<td>2 (3)</td>
<td>—</td>
</tr>
<tr>
<td>Ideal weight [n (%)]</td>
<td>176 (73)</td>
<td>43 (61)</td>
<td>—</td>
</tr>
<tr>
<td>Overweight [n (%)]</td>
<td>51 (21)</td>
<td>21 (30)</td>
<td>—</td>
</tr>
<tr>
<td>Obese [n (%)]</td>
<td>10 (4)</td>
<td>4 (5)</td>
<td>—</td>
</tr>
<tr>
<td>Reference AUC (mmol·min⁻¹L⁻¹)</td>
<td>206 ± 5</td>
<td>258 ± 13</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Reference CV (%)</td>
<td>22.3 ± 1.0</td>
<td>23.1 ± 2.9</td>
<td>NS</td>
</tr>
<tr>
<td>Glycemic index (%)</td>
<td>53.1 ± 1.2</td>
<td>53.8 ± 2.4</td>
<td>NS</td>
</tr>
</tbody>
</table>

² 26 Southeast-Asian, 21 East-Asian, 16 South-Asian, 3 mixed African/South-Asian, 3 African, and 1 Middle-Eastern.

³ x ± SEM (all such values).

Mean area under the curve for reference food; values for white bread were adjusted by dividing by 0.71.

Mean glycemic index of 2 test foods for 287 subjects who tested both foods.
150 subjects at these laboratories, 30 reported taking medications, the most common of which was the oral contraceptive pill (n = 18). Other medications were angiotensin-converting enzyme inhibitors (n = 3), nonsteroidal anti-inflammatory agents (n = 3), thyroid replacement therapy (n = 2), and one subject each for aspirin, β-blocker, anti-histamine, antibiotic, and a triptan for migraines. Information about the subjects’ smoking habits was provided by 12 laboratories, and of the 136 subjects involved, 3 were smokers. At 26 of the 28 participating laboratories, all subjects (n = 275) tested both test foods. In the 27th laboratory, 18 subjects were involved, 4 of whom tested both test foods. In the 28th laboratory, 19 subjects were involved, 3 of whom tested both foods.

Some or all of the AUC values reported by 14 (54%) of the 26 laboratories that reported AUC values to the central laboratory were considered to be calculated or reported incorrectly (Figure 1A). The limits of agreement for 10 of these laboratories were relatively small, ranging from ±5 to ±21 mmol·min/L; but for 4 of the laboratories, the limits of agreement were >±50 mmol·min/L. One of the latter laboratories appeared to have reported total AUC values instead of incremental AUC. One laboratory did not report AUC or GI values. A total of 52 mean GI values were reported to the central laboratory (2 foods tested in 26 laboratories); of these, 23 (44%) differed from those calculated by the central laboratory by >1 (Figure 1B). Ten (19%) of the mean GI values differed from those calculated by the central laboratory by >5, with at least one difference of >5 being reported by 7 (27%) of the laboratories. These differences were due, at least in part, to the use of different exclusion criteria by most laboratories; 25 laboratories did not exclude any outliers when reporting mean GI values, 2 laboratories excluded values >2 SDs from the mean, and one laboratory excluded values <10.1 and >99.9. When outliers were not excluded by the central laboratory, mean GI values calculated by the central laboratory differed by >5 from those reported by the laboratories for 9 foods reported by 4 different laboratories. One of these laboratories was the one that excluded values <10.1 and >99.9, but the other 3 laboratories did not exclude any values. Thus, the difference in reported GI was not due to differences in exclusion criteria for these 3 laboratories.

Individual values for mean AUC (mean of test and reference foods) varied from 23 to 413 mmol·min/L. By stepwise multiple regression analysis, the individual values for AUC were not related to sex or age but were significantly related to ethnicity (lower in white subjects than in others, P < 0.001) and BMI (P = 0.031; positive relation), with the combined r² = 0.095. By stepwise multiple regression analysis, the individual values for GI were negatively related to refCV (r² = 0.017, P = 0.033), but were not related to sex, ethnicity, age, or BMI with the use of either simple (univariate) or multiple regression analysis. GI values of the individual subjects were positively related to refCV (r = 0.143, P = 0.020), but were not related to sex, ethnicity, age, BMI, or mean AUC with the use of either univariate or multiple regression analysis.

FIGURE 1. A: Differences between reported and centrally calculated individual area under the curve (AUC) values. Values are mean ± SD. The scale for values represented by open squares (right y axis) is 10 times that for values represented by closed circles (left y axis). B: The difference between reported and centrally calculated mean glycemic index (GI) values for 2 foods plotted against the mean of the reported and centrally calculated values. The dotted lines represent the differences in GI that are considered to be large, which were arbitrarily chosen to be >5.

FIGURE 2. Mean blood glucose responses elicited by the reference food, cheese puffs (●), and fruit leather (△) in 24 laboratories (n = 271 subjects) that used glucose (○) as the reference food (left) and in 4 laboratories (n = 70 subjects) that used white bread (□) as the reference food.
The blood glucose responses of the 2 test foods were similar in the 24 laboratories that used glucose as the reference food compared with the 4 that used white bread (Figure 2). The mean AUC for the reference food (values for white bread adjusted to the glucose scale) differed significantly between laboratories (F(25, 301) = 4.39, P < 0.001; one subject with an AUC of 666 mmol·min/L was excluded as an outlier; Figure 3A). In addition, the mean within-subject CV for the repeated reference food tests (refCV) differed in the different laboratories (F(25, 299) = 2.42, P < 0.001, Figure 3B; 2 subjects with CV > 141% were excluded as outliers). The mean adjusted AUC and refCV values for the 4 laboratories that used white bread as the reference were not significantly different from those laboratories that used glucose (Figure 3).

GI values for cheese puffs were available in 300 subjects, of whom 5 were outliers (>2 SDs above the laboratory mean) and for fruit leather in 298 subjects of whom 8 were >2 SDs above the mean. Thus, of the 598 individual GI values, 13 (2.2%) were considered outliers and were excluded. By ANOVA, there was a significant main effect of food (F(1, 237) = 476, P < 0.001), with the mean GI of fruit leather being significantly less than that for cheese puffs (Figures 4, A and B). The main effect of laboratory was not significant (F(25, 249) = 1.529, P = 0.056), and there was no significant food × laboratory interaction (P = 0.43). The mean ± SD of the 28 laboratory mean GI values for cheese puffs and fruit leather, respectively, were 74.3 ± 10.5 (Figure 4A) and 33.2 ± 7.2 (Figure 4B). The SDs of the GI values in the 28 laboratories varied from 14 to 48 for cheese puffs (mean±SD, 27.5 ± 9.0; Figure 4C) and from 9 to 25 for fruit leather (14.7 ± 4.2, Figure 4D).

Methodologic variables related to the test meal included the type of reference food used, the number of reference food tests done by each subject, and the nature of the drink given with the test meal. By ANOVA, the variation in these factors was not associated with any significant differences in adjusted reference AUC, refCV, mean GI, or SD of GI (Table 3).

The laboratories varied in the methods used for blood sampling and glucose analysis. The laboratories (n = 4) that took 2 fasting blood samples and used the average measure to calculate GI tended to have a lower mean refCV and mean GI (not significant) and a significantly lower SD (Table 4). Duplicate measurement of glucose also tended to be associated with reduced variation, although the difference was not significant. The 11 laboratories that did any duplication (duplicate blood samples, duplicate measurements, or both) had a significantly lower SD than did the 17 laboratories that did not. Use of an enzymatic method to measure glucose tended to be associated with higher refCV and SD of GI than the other analytic methods, although the differences were not statistically significant (Table 4).

FIGURE 3. A: Incremental area under the glycemic response curve (AUC) elicited by the reference food in different labs. B: Coefficient of variation (CV, where CV = 100 × SD/mean) of the AUC values elicited by the reference food (tested 2–3 times by each subject) in different laboratories. Values are the mean ± SEM for glucose (●, n = 22) and white bread (○, n = 4; AUC values for white bread were divided by 0.71). Open diamonds represent the mean ± SEM for the 26 laboratories (reference CV could only be calculated for the 26 laboratories that tested the reference food more than once in each subject).

FIGURE 4. Mean and SD of glycemic index (GI) values from the 28 laboratories for cheese puffs (A and C, respectively) and fruit leather (B and D, respectively). Closed circles are for laboratories using glucose as the reference food; open circles are for laboratories using white bread as the reference food, and the open diamond represents the mean and SD of the values from the 28 laboratories.
There was wide variation in the restrictions placed on the subjects’ diets and activities before the test in different laboratories. Allowing the subjects to do usual, but not unusual, physical activities (n = 5 laboratories) for 24 h before the test was associated with significantly lower mean GI and a tendency toward reduced variation than that reported by the 10 laboratories that did not place any restrictions on the subjects’ activities before the test. The results for the 13 laboratories that did not allow any vigorous physical activity whatsoever were intermediate (Table 5). Not allowing subjects to consume alcoholic beverages for 24 h before the test and restricting the type of dinner consumed or providing a standardized dinner were associated with nonsignificant reductions in the mean and SD of GI values (Table 5). The length of the fasting time did not appear to have any significant effect.

By univariate regression analysis, the mean GI value of the laboratories was significantly related to 2 factors: laboratory mean refCV (r = 0.556, P = 0.003; Figure 5A) and the type of restriction placed on exercise before the test (exercise number: 0 = no restriction, 1 = no activity at all, 2 = no unusual activity; r = −0.495, P = 0.010). By multiple regression analysis, refCV (P = 0.006) and exercise number (P = 0.017) together explained 46% of the variation in mean GI (i.e., the coefficient of determination: $r^2 = 0.463, r = 0.681, P < 0.001$). However, the combination of refCV ($P < 0.001$) and the restriction placed on

---

**TABLE 3**

<table>
<thead>
<tr>
<th>Reference food</th>
<th>Reference AUC</th>
<th>Reference CV</th>
<th>Mean GI</th>
<th>SD of GI</th>
<th>DV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mmol · min/L</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>Glucose (n = 24)</td>
<td>213 ± 12</td>
<td>22.0 ± 1.2</td>
<td>58.5 ± 1.6</td>
<td>22.1 ± 1.0</td>
<td>1</td>
</tr>
<tr>
<td>White bread (n = 4)</td>
<td>230 ± 8</td>
<td>26.2 ± 5.0</td>
<td>62.9 ± 5.6</td>
<td>26.1 ± 6.8</td>
<td>2</td>
</tr>
</tbody>
</table>

---

**TABLE 4**

<table>
<thead>
<tr>
<th>Brain sampling and glucose measurement methods on outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference AUC</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Duplicate fasting blood sample</td>
</tr>
<tr>
<td>No (n = 24)</td>
</tr>
<tr>
<td>Yes (n = 4)</td>
</tr>
<tr>
<td>Duplicate measurement of glucose</td>
</tr>
<tr>
<td>None (n = 19)</td>
</tr>
<tr>
<td>Fasting only (n = 2)</td>
</tr>
<tr>
<td>All blood samples (n = 6)</td>
</tr>
<tr>
<td>Any duplication</td>
</tr>
<tr>
<td>No (n = 17)</td>
</tr>
<tr>
<td>Yes (n = 11)</td>
</tr>
</tbody>
</table>

---

1 The number of laboratories is indicated by n. Shown is the incremental area under the curve (AUC) for the reference food (glucose or white bread); values for white bread were adjusted by dividing by 0.71. Reference CV is the CV (100 × SD/SD) of the reference AUC values for each subject. DV is the value of the dummy variable assigned to the method variable for step-wise multiple regression analysis. GI, glycemic index. Means not sharing the same superscript letter are significantly different, P < 0.05.

2 Enzymatic tests included Analox GM9D Glucose Direct analyser (Waltsbam, MA) (n = 1), Hemocue is Hemocue 201+ analyzer, Hemocue AB (Ångelholm, Sweden). Glucometers used included Glucotrend (Roche Diagnostics, Basel, Switzerland) (n = 1), Lifescan Surestep (Milpitas, CA) (n = 1), Ascensia (Bayer Diagnostics, Tarrytown, NY) (n = 2), and Medisense (Abbott Laboratories, Abbott Park, IL) (n = 1). One laboratory, which measured fasting glucose by meter (type not specified) and postprandial glucose by using YSI, was placed into the Glucometer category. YSI is Model 2300 STAT glucose analyzer (YSI Inc, Yellow Springs, OH). The dummy variable used for YSI was 1 for correlations with mean GI and 3 for correlations with SD of GI.
GLYCEMIC INDEX INTERLABORATORY STUDY

TABLE 5
Effect of subject preparation variables on outcomes

<table>
<thead>
<tr>
<th>Reference AUC</th>
<th>Reference CV</th>
<th>Mean GI</th>
<th>SD of GI</th>
<th>DV</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmol · min/L</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
</tbody>
</table>

Vigorous activity

<table>
<thead>
<tr>
<th>No restriction (n = 10)</th>
<th>229 ± 16</th>
<th>24.2 ± 2.3</th>
<th>58.5 ± 2.2*</th>
<th>26.0 ± 2.0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>No unusual (n = 5)</td>
<td>185 ± 23</td>
<td>20.5 ± 2.3</td>
<td>50.1 ± 1.2b</td>
<td>19.4 ± 1.9</td>
<td>2</td>
</tr>
<tr>
<td>None at all (n = 13)</td>
<td>216 ± 15</td>
<td>22.2 ± 1.8</td>
<td>51.4 ± 2.0ab</td>
<td>21.4 ± 1.8</td>
<td>1</td>
</tr>
</tbody>
</table>

Alcohol consumption

<table>
<thead>
<tr>
<th>Not specified (n = 5)</th>
<th>217 ± 13</th>
<th>21.2 ± 2.7</th>
<th>60.0 ± 3.0</th>
<th>22.5 ± 0.9</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>None for 12 h (n = 9)</td>
<td>208 ± 18</td>
<td>20.4 ± 2.2</td>
<td>53.6 ± 2.1</td>
<td>22.6 ± 1.8</td>
<td>1</td>
</tr>
<tr>
<td>None for ≥24 h (n = 14)</td>
<td>219 ± 16</td>
<td>24.5 ± 1.7</td>
<td>50.5 ± 1.5</td>
<td>18.0 ± 1.6</td>
<td>2</td>
</tr>
</tbody>
</table>

Dinner the night before

<table>
<thead>
<tr>
<th>No restriction (n = 11)</th>
<th>215 ± 9</th>
<th>24.5 ± 2.3</th>
<th>55.8 ± 2.6</th>
<th>24.2 ± 1.9</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrictions (n = 12)</td>
<td>214 ± 19</td>
<td>21.9 ± 1.5</td>
<td>52.4 ± 1.7</td>
<td>19.2 ± 1.3</td>
<td>1</td>
</tr>
<tr>
<td>Meal provided (n = 5)</td>
<td>217 ± 28</td>
<td>20.1 ± 3.1</td>
<td>52.3 ± 3.4</td>
<td>18.9 ± 2.8</td>
<td>2</td>
</tr>
</tbody>
</table>

Length of fasting

<table>
<thead>
<tr>
<th>Fixed time (n = 18)</th>
<th>208 ± 15</th>
<th>23.6 ± 1.4</th>
<th>54.1 ± 1.9</th>
<th>21.7 ± 1.3</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>None or range (n = 10)</td>
<td>229 ± 5</td>
<td>20.9 ± 2.4</td>
<td>53.1 ± 2.0</td>
<td>20.0 ± 2.2</td>
<td>1</td>
</tr>
</tbody>
</table>

1 The number of laboratories is indicated by n. Shown is the incremental area under the curve (AUC) for the reference food (glucose or white bread); values for white bread were adjusted by dividing by 0.71. Reference CV is the CV (100×SD/mean) of the reference AUC values for each subject. DV is the value of the dummy variable assigned to the method variable for step-wise multiple regression analysis. GI, glyemic index. Means not sharing the same superscript letter are significantly different, P < 0.05.

2 No restriction includes laboratories stating there was no restriction (n = 6) and laboratories not specifying (n = 4). None at all includes laboratories that did not allow subjects to undertake any vigorous activity for 3 d (n = 1), 24 h (n = 8), 18 h (n = 1), or 12 h (n = 3) before the test.

3 ≥24 h includes laboratories that advised subjects to consume no alcohol for 24 h (n = 12) or 48 h (n = 1) and laboratory that excluded subjects who ever used alcohol.

4 Restrictions includes laboratories allowing subjects to self-select the dinner but advising them to always have the same dinner (n = 4), to avoid high-fiber foods (n = 3), to avoid legumes (n = 2), to have a high-carbohydrate, low-fat meal (n = 2), and to have a high-carbohydrate, high-GI meal (n = 1). The types of meals provided to subjects included solid foods (n = 3), liquid test meals (n = 1), and white bread and water (n = 1).

5 Fixed time includes laboratories that advised subjects to fast for 8 h (n = 1), 10 h (n = 9), or 12 h (n = 8). None or range includes laboratories that did not specify the length of fasting time (n = 3) or advised subjects to fast for 10–12 h (n = 6) or 10–14 h (n = 1).

6 Different laboratories calculate the AUC and handle outlying values. These issues will have to be addressed for regulatory purposes. When centrally calculated GI values were considered, the reliability of the GI method reported here was similar to that in a previous study (32) in which the average between-laboratory SD of GI values was 9. However, in the present study, we were also able to see whether the characteristics of the subjects studied or subtle variations in the methods used were associated with the accuracy (reflected by the mean value) and precision (reflected by the SD) of the GI results obtained by different laboratories.

7 Consistent with previous knowledge (37–39), the glycemic responses (ie, AUCs) of individual subjects were associated with ethnicity and BMI. However, ethnicity, sex, age, and BMI were not related to the within-individual variation in glycemic responses (ie, refCV) nor, most importantly, to GI values in the individual subjects. This is consistent with previous studies suggesting that, when measured by using appropriate methods, GI is the same in different subjects (31, 32) and therefore is a property of the food and not of the subject in whom it is measured. The implications of this include not only that GI can be measured validly in most subjects but also that the results apply to most of the healthy population.

The distribution of GI values in individual subjects is skewed to higher values because that is a mathematical property of the ratio of 2 independently variable measures (40). Repeating AUC measurements can reduce skewness, which is most cost-effectively achieved by repeating the reference food (32, 41). Nevertheless, very high individual values still occur, and they
increase the resulting mean and SD. Thus, it has been recommended that outliers should be excluded (31). Values >2 SDs above the mean are often considered to be outliers, a cutoff that excludes 2.5% of normally distributed values. Because GI values are not normally distributed, a cutoff of 2 SDs may not be considered conservative enough if it results in >2.5% of the values being excluded. However, our results show that, despite a skewed distribution, only 2.2% of individual values were >2 SDs above the mean (95% CI: 1.3–3.4%). Thus, >2 SDs above the mean appears to be an appropriate definition of GI outliers. Only 2 (0.3%) values were >2 SDs below the mean, and, because GI values are skewed to the right, we did not consider it appropriate or necessary to exclude these values.

Within-individual variation in glycemic responses (laboratory mean refCV) was positively related to the mean and the SD of GI values. The implication of this is that a high refCV leads to bias and imprecision of the resulting GI values. Previous work suggests that refCV is affected by the blood sampling schedule, method of calculating AUC (42), capillary versus venous blood sampling (32, 41), and the restrictions placed on the subjects’ diet and activity the day before testing (43). Here, however, the only factor significantly related to refCV was that subjects with a low average AUC tended to have a high refCV. This could be because the proportion of total variation due to analytic variation becomes larger as the AUC becomes smaller (44). Overweight subjects or those with a family history of diabetes are sometimes excluded from GI testing, presumably to reduce variation. Ironically, however, these exclusion criteria would have the effect of biasing the resulting GI values toward being too high and making them imprecise, because they would select for subjects with a low AUC and, hence, a high refCV. The present results suggest that mean GI is not correlated with refCV when mean refCV is <28% (Figure 5A), and the SD of GI is not correlated with refCV when mean refCV is <26% (Figure 5C).

The only factors related to mean GI, other than refCV, were that restrictions on exercise and alcohol consumption before the test were associated with lower mean GI. This unexpected result is hard to explain. Asking subjects to avoid alcohol and exercise doesn’t guarantee that they will avoid them, and giving no advice doesn’t mean that the subjects will be exposed to them. Prior alcohol consumption and physical activity may reduce glycemic responses by reducing hepatic glucose output (45, 46) and improving insulin sensitivity (47–49), respectively. However, to affect GI, the glycemic response elicited by the reference food

![Figure 5](https://academic.oup.com/ajcn/article-abstract/87/1/247S/4633409)
would have to be affected to a different extent than that after the test food. Most laboratories used glucose as the reference food. The cheese puffs contained more protein and fat than the glucose did, and both test foods contained more dietary fiber than the glucose did. The extent to which protein and fat reduce glycemic responses may depend on subjects’ fasting plasma insulin, waist circumference, fiber intake (50, 51), insulin sensitivity (52), and possibly fat intake (53). It is not known whether the effects of alcohol and exercise on acute glycemic responses interact with those of other dietary factors.

We expected methodologic variables to influence the within-laboratory SD of GI values; however, the nature of some of the associations found was unexpected. In previous studies, the use of 2 or 3 reference tests resulted in a lower SD than the use of only 1 (31, 41). In the present study, we could find no evidence to justify doing 3 rather than 2 tests, because the difference was small and not significant. Most laboratories allowed the subjects to drink only water with the test meals, presumably to reduce confounding factors. Unexpectedly, however, even though caffeine acutely increases glycemic responses (54), allowing drinks of coffee and tea tended to be associated with a lower SD of GI than drinking only water (NS; Table 3).

Taking 2 fasting blood samples, doing duplicate measurements of glucose, and using the YSI or glucometer to measure glucose compared with the other methods used were associated with a lower intralaboratory SD of GI. These results differ from those of recent studies that compared some of these factors under more controlled conditions. In a head-to-head comparison of glucose analytic methods, GI values calculated from glucose concentrations measured by YSI were more precise than those measured in the same blood samples by using the One Touch Ultra glucometer (55). However, the precision of different glucose meters varies (56), and we did not measure the precision of the glucose analysis methods used in each laboratory. Recently, it was shown that analyzing glucose 2 times in a single fasting sample reduced the SD of GI values more than did using the average of 2 different fasting samples (57). This was suggested to be due to the fact that, in that study, the variability of the glucose analytic method (CV < 2%) was less than the minute-to-minute variation in fasting glucose (CV ≈3%) (57); with a less precise analytic method, it may be advantageous to take 2 fasting blood samples.

Our results show that restricting the type of dinner the subjects consumed the night before the test or providing a standard meal was associated with significantly lower SD of GI values. This is presumably because the nature of the meal consumed the night before can carry over to influence the glycemic response the next morning (58, 59). Providing a standardized meal to subjects increases the cost of doing GI testing; our results suggested that simply advising subjects to avoid certain types of foods is almost as good and may be more cost-effective.

The results of this study should not be used to define the required methods for GI testing for several reasons. Associations do not prove causality; thus, our results should be used to develop hypotheses for further testing under controlled conditions. The costs versus benefits need to be considered. For example, the same reduction in the margin of error achieved by providing all subjects with a standard meal the night before every test may be able to be achieved more economically by adding 1 or 2 more subjects. Also, it may not be necessary to do all of the things associated with reduced SD to obtain satisfactory results.

**SUMMARY OF RECOMMENDATIONS**

**Factors affecting the accuracy of GI**

1) Calculating AUC; errors are common (50% of laboratories), and standardization is required.

2) Outliers: exclude values >2 SDs from the mean (within laboratories).

3) Within-individual variability of AUC; mean refCV for the subject group should be <30%.

**Factors affecting the precision of GI**

4) Glucose analysis: the use of a precise analytic method, duplicating blood samples, or duplicating glucose analyses may be beneficial; specific recommendations may depend on the analytic method used.

5) Subject preparation: moderate restrictions (eg, asking subjects to have a normal meal the night before and to avoid unusual exercise) may be beneficial, but expensive or rigorous restrictions (eg, providing subjects with a standardized meal the night before and prohibiting any physical activity before the test) may have little or no additional effect.

**Factors not necessary to control**

6) Subject characteristics: no need to restrict the age, sex, BMI, or ethnicity of the subjects.

7) Test meal: there may be no need to avoid coffee or tea as the drink with the test meal.

**CONCLUSIONS**

The between-laboratory SD of GI values is ≈9 with no significant difference in mean GI between laboratories. Standardized data analysis and low within-subject variation (refCV < 30%) are required for accuracy. The results suggest that common misconceptions exist about which factors do and do not need to be controlled to improve precision. Controlled studies and cost-benefit analyses are needed to optimize GI methodology.

The contributions of the authors were as follows. TMSW: had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The following letters identify the types of contributions and are listed for each author. A: conception and design of overall study; B: drafting and circulation of protocol, provision of central foods, central data analysis; C: critical review of protocol (at each center; design of procedures, acquisition of data, and calculation of results); D: Drafting and revision of the manuscript; E: critical revision of the manuscript for important intellectual content; F: obtained funding. TMS Wolever: A, B, C, D, F; JC Brand-Miller: A, C, E, F; J Abernethy: C, E, F; A Astrup: C, E, F; P Atkinson: C, E; M Axelsen: C, E, F; Björk: C, E, F; Brighenti: C, E, F; Brown: C, E; A Brynes: C, E; MC Casaniglia: C, E, F; M Caraubiel: C, E, F; L Dahlqvist: C, E; E Delport: C, E, F; GS Denyer: C, E; D Erba: C, E; G Frost: C, E, F; Y Granfeldt: C, E; S Hampton: C, E, F; VA Hart: C, E; K Häätönen: C, E; CJ Henry: C, E, F; S Hertzler: C, E, F; S Hull: C, E; J Jerling: C, E, F; KL Johnston: C, E, F; H Lightowler: C, E; N Mann: C, E, F; L Morgan: C, E, F; F Panlasigui: C, E, F; Pekelman: C, E, F; T Perry: C, E, F; AH Pfeiffer: C, E, F; M Pieters: C, E; DD Ramdath: C, E, F; RT Ramsingh: C, E, F; SD Robert: C, E, F; C Robinson: C, E; Sarkkinen: C, E, F; A Scazzina: C, E; DCD Sison: C, E; B Sloth: C, E; J Staniforth: C, E, F; N Tapola: C, E, L Valsta: C, E, F; I Verkkoijen: C, E; MO Weickert: C, E; A Weseler: C, E, F; P Wilke: C, E, J Zhang: C, E, F.

Author statements of conflicts of interest are as follows. TMS Wolever: I am president and part-owner of Glycemic Index Laboratories Inc, a contract research organization and president and part-owner of Glycemic Index
Testing Inc, a corporation that provides services related to the measurement of the glycemic index of foods. I received grant and research support from Cargill Inc and ILSI Europe, was a consultant for the US Potato Board, and received honoraria for consulting and speaking from the Dutch Sugar Bureau and Mars Inc. I am co-author of a range of popular books on the glycemic index under the general title of *The Glucose Revolution: Authoritative Guide to the Glycemic Index*, published by Marlowe & Co, NY. I am the author of a scientific book titled *The Glycemic Index: A Physiologic Classification of Dietary Carbohydrate*, published by CABI, UK. Jennie C Brand-Miller serves on the board of directors of Glycemic Index Limited, a not-for-profit company that administers the Glycemic Index Symbol food labeling program in Australia (www.gisymbol.com.au). She is also a director of a not-for-profit glycemic index testing service at the University of Sydney (www.glycemicindex.com). She is a co-author on a series of books under the general title *The New Glucose Revolution* (published by Marlowe and Co in North America), which explains the theory and practice of the glycemic index to the lay public. John Aherne is president of AMK Research, Inc, a research company that performs GI testing. Arne Astrup is receiving research funding to dietary intervention studies from >100 food producers and companies and has received speakers’ honoraria (Danish Meat Council, Arla Foods, Danish Dairy Board, and Unilever) and fees for participation in advisory boards (Arla Nutrition Advisory Board, European Almond Advisory Board, Communications and Scientific Advisory Board of the Global Dairy Platform, Proctor & Gamble, Novartis, and Unilever) and is a medical advisor for Weight Watcher Denmark. Furio Brighenti received research funding and consultancy fees from Barilla R&G F.Ili. SpA through ParmaTecninova Srl, a company partly owned by the University of Parma, and at the time of the study was a member of the following scientific boards: Beneo (a brand of Orafti group, Belgium) and Soremarcet (a company owned by Ferrero SpA).

Elizabeth Delport is one of the executive members of the GI Foundation of SA, a contract research organization, and is a co-author on popular books on GI. Shelagh Hampton and Linda Morgan are members of a university team that conducts glycemic index testing for Surrey University. Valerie Hart is employed by Reading Scientific Services, Ltd. C Jeya Henry is a member of a university team that conducts glycemic index testing for Oxford Brooks University. Sarah Hull is employed by Leatherhead Food International. Kelly L. Johnston was employed by Leatherhead Food International at the time of the study. Helen Lightowler is a member of a university team that conducts glycemic index testing for Oxford Brooks University.

REFERENCES

GLYCEMIC INDEX INTERLABORATORY STUDY


42. Wolever TMS. Effect of blood sampling schedule and method calculating the area under the curve on validity and precision of glycemic index values. Br J Nutr 2004;91:295–300.


