

The reproducibility of subjective appetite scores

BY ANNE RABEN¹, ANNA TAGLIABUE² AND ARNE ASTRUP¹

¹ Research Department of Human Nutrition, Centre for Food Research, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark

² Department of Human Nutrition, University of Pavia, Italy

(Received 28 March 1994 – Revised 13 June 1994 – Accepted 24 August 1994)

Although subjective appetite scores are widely used, studies on the reproducibility of this method are scarce. In the present study nine healthy, normal weight, young men recorded their subjective appetite sensations before and during 5 h after two different test meals A and B. The subjects tested each meal twice and in randomized order. Visual analogue scale (VAS) scores, 10 cm in length, were used to assess hunger, satiety, fullness, prospective food consumption and palatability of the meals. Plasma glucose and lactate concentrations were determined concomitantly. The repeatability was investigated for fasting values, Δ -mean 5 h and mean 5 h values, Δ -peak/nadir and peak/nadir values. Although the profiles of the postprandial responses were similar, the coefficients of repeatability (CR = 2SD) on the mean differences were large, ranging from 2.86 to 5.24 cm for fasting scores, 1.36 to 1.88 cm for mean scores, 2.98 to 5.42 cm for Δ -mean scores, and 3.16 to 6.44 cm for peak and Δ -peak scores. For palatability ratings the CR values varied more, ranging from 2.38 (taste) to 8.70 cm (aftertaste). Part of the difference in satiety ratings could be explained by the differences in palatability ratings. However, the low reproducibility may also be caused by a conditioned satiation or hunger due to the subjects' prior experience of the meals and therefore not just be a reflection of random noise. It is likely, however, that the variation in appetite ratings is due both to methodological day-to-day variation and to biological day-to-day variation in subjective appetite sensations.

Appetite: Hunger: Satiety: Visual analogue scale

Subjective hunger and satiety may be influenced by a number of different factors, including physiological and psychological variables (Blundell, 1990). The latter may be particularly important when evaluating the impact of different meals and diets on appetite sensations, as psychological factors may interfere with the truly physiological effects of the meal and thus bias the results. Moreover, subjective appetite sensations on the test day may be influenced by environmental factors irrelevant to the study, i.e. climate and season. This may affect both the physiological and psychological factors.

In order to assess subjective appetite sensations, visual analogue scale (VAS) scores are most often used. These can be composed as lines of varying length (9, 10, 15, 20 points or cm) with words anchored at each end expressing the most positive or negative sensation (i.e. I have never been more hungry/I am not hungry at all; Haber *et al.* 1977; Hill *et al.* 1984; Burley *et al.* 1993). Subjects make a point on the line corresponding to their hunger sensation. The distance from the end is measured and often basal ratings are deducted from the post-meal ratings.

The reproducibility of VAS scores has been quite well studied in other research areas. In a study on rheumatic patients on the reproducibility along a 10 cm vertical VAS it was concluded that the error involved in the use of VAS scores was even more complex than previously thought (Dixon & Bird, 1981). Within pain research, caution in the use of VAS

scores has been recommended by some (Lahuerta *et al.* 1983), while others have found VAS scores a satisfactory and more reliable method compared with other commonly used methods (Joyce *et al.* 1975; Ohnhaus & Adler, 1975). Likewise, VAS scoring used to estimate the sensations of dyspnoea during exercise in dyspnoeic patients was found to provide both a reproducible and sensitive estimation method (Loiseau *et al.* 1990).

The results from these studies cannot, however, be extrapolated to appetite research. Here it appears that only very few studies have investigated the reproducibility of VAS scores. Instead, some studies have investigated the correlation between hunger ratings and subsequent energy intake, i.e. the validity of the appetite ratings. Many of these have failed to demonstrate such a relationship (Pi-Sunyer *et al.* 1982; Rogers *et al.* 1988; Rolls *et al.* 1988; Mattes, 1990), while others have found that there was a connection (Silverstone & Stunkard, 1968; Spiegel *et al.* 1987; Blundell & Hill, 1988). Together this implies that hunger sensations under certain, but not all, circumstances can be used as a reliable measure of energy intake.

In a study by Leathwood & Pollet (1988) six subjects were given six different test meals in triplicate. They subsequently rated their hunger sensations on a 10-point scale. The pooled within-subject, within-treatment standard deviations for satiety, fullness, hunger and 'gourmandise' were 2.09, 2.06, 2.10 and 2.10 units. According to the authors this showed that ratings, within a given treatment, were consistent. However, they also found that there was an increase in within-subject variability in ratings over time indicating problems in reproducing the scores (Leathwood & Pollet, 1988).

In a recent study by Lappalainen *et al.* (1993) eight subjects were given a breakfast with or without water in triplicate after which they rated their hunger sensations on a 100 mm VAS. From a paired rank-sum test it was concluded that there was no difference in mean values for hunger, satiety and desire to eat after the three identical test meals and the results were therefore pooled. This is, however, not the same as testing the repeatability of scores (Bland & Altman, 1986). In a previous study by Bolton *et al.* (1981) two different scoring systems were used, a scale score and a verbal score system. A high correlation coefficient led the authors to conclude that the validity of the methods was thereby shown. This is, however, also not a proof of a method's reproducibility (Bland & Altman, 1986).

Silverstone & Goodall (1986) have looked at reproducibility of VAS scores by use of a pharmacological approach. They argued that if an effective anorectic drug modifies hunger ratings in one population, it should, if the measurement is reliable, affect hunger ratings similarly in another population. They did indeed find this to be the case in an examination of the effects of the drug tiflorex. They found that the anorectic effect of this compound over an 8 h period in a first study was closely mirrored by the effect observed in a second study (Silverstone & Goodall, 1986). These findings therefore support the validity of subjective appetite scores.

Although widely used, the reproducibility of appetite scores in meal-test studies has still been little investigated. In the present study appetite and palatability ratings were assessed twice after two different test meals. Moreover, plasma glucose and lactate concentrations were measured simultaneously in order to get a measure of the impact of the meals on these physiological variables.

METHODS

The study was part of a larger multi-centre study within the EURESTA (European FLAIR-concerted action on resistant starch) framework. The present data derive from exactly the same experiment conducted at two centres; the Research Department of Human Nutrition, The Royal Veterinary and Agricultural University, Denmark and the Department of Human Nutrition, University of Pavia, Italy.

Subjects

Nine healthy male subjects, 23.7 (SEM 1.3) years of age, normal weight (body mass index 22.4 (SEM 0.7) kg/m²), non-smokers, not elite athletes, and with no history of obesity or diabetes, completed the study. The subjects' energy needs during the study were determined using World Health Organization tables according to age, weight, height and sex (Food and Agriculture Organization/World Health Organization/United Nations University, 1985). The multiplication factor 1.78 was used to account for medium physical activity level of the subjects. The body composition of the subjects was estimated by the bioimpedance method by an Animeter (HTS-Engineering Inc., Odense, Denmark) in Denmark and BIA 109 (Ryl System, Detroit, USA) in Italy. Fat-free mass was calculated using the equation of Deurenberg *et al.* (1991). The study was approved by the Municipal Ethical Committees of Copenhagen and Frederiksberg, Denmark and Pavia, Italy to be in accordance with the Helsinki-II declaration. All subjects gave written consent after the experimental procedure had been explained to them.

Diets

Two different test meals were used. One meal consisted of 50 g raw potato starch (13.6 g digestible starch) (A1 and A2), the other of 50 g pregelatinized potato starch (46.5 g digestible starch) (B1 and B2). For both the A and B meals the starch was mixed into 500 ml diluted, artificially sweetened fruit syrup (8.4 g fruit sugars). The syrup was based on apple, grape, redcurrant, elderberry, blackcurrant and cherry and commercially purchased from Irma, Denmark. The starches were produced and supplied by l'Institut National de la Recherche Agronomique (INRA), Nantes, France. Meal A was a drinkable mixture, while meal B was a porridge to be eaten by spoon. Total available energy content of the meals, not taking large-intestinal starch fermentation into consideration (Mathers, 1991), was calculated to be 367 kJ for A and 917 kJ for B.

Each test day was preceded by 3 d on an identical carbohydrate-rich diet (60% energy (E%) carbohydrate, 28 E% fat, 12 E% protein, 3.5 g dietary fibre/MJ), prepared at the department from national food items according to each subjects' individual energy requirements, and adjusted to the nearest 0.5 MJ. The subjects were instructed to adhere strictly to the diet. If they could not consume all the food, they had to bring the leftovers to the department for weighing and registration. The same food was deducted from the diet during the following pre-experimental periods. The subjects were instructed to abstain from physical activity for 2 d before the test days. Together with the standard diet this should ensure equally filled glycogen stores and similar macronutrient balance on the two test days (Costill, 1988). The computer database of foods from the National Food Agency of Denmark (Dankost) was used in the calculations of energy and nutrient composition of the test diets.

Experimental protocol

The four test meals were given in different order for the nine subjects on separate days with at least 1 week and no more than 3 weeks separating the test days. On the test day the subjects arrived at the institute with a minimum of physical activity using car, bus or train after having fasted for 12 h from the evening before. After voiding and weighing (to the nearest 100 g), bioimpedance was measured by an Animeter (HTS-Engineering Inc.). The subjects then rested in the supine position on a bed covered with an antidecubitus mattress and with a slight elevation of the head. A Venflon catheter (Viggo, Gothenborg, Sweden) was inserted in an antecubital arm vein. After 10 min rest a fasting blood sample was taken and after a further 60 min rest a second fasting blood sample was taken. In between, basal energy expenditure was measured using a ventilated hood system. The test meal was then

served and eaten within 10 min. Exactly the same time was spent on the four meals for each individual subject. Blood samples were taken after 15, 30, 45, 60, 75, 90, 105, 120, 150, 180, 210, 240 and 300 min. During the postprandial measurements energy expenditure was measured and the subjects were allowed to watch light entertainment movies, and to have a break at 12.00 and 14.00 hours of a maximum of 5 min. During the break the subjects could sit, walk quietly or go to the toilet. The exact hour, the time spent and the type of activity during the break was noted and repeated on the following test days. Water consumption during the test day was allowed, but the total amount consumed was noted and repeated on the next three test days.

Immediately before and every 30 min after the meal, questionnaires to assess hunger, fullness, desire to eat and prospective food consumption were filled out by each subject. Ratings were made on 10 cm VAS with words anchored at each end, expressing the most positive (e.g. good, pleasant) or the most negative rating (e.g. bad, unpleasant) (Hill *et al.* 1984). Immediately after the test meals the palatability, taste, after-taste, texture and visual appeal of the two test meals were recorded by the subjects using VAS scores (Hill *et al.* 1984).

Laboratory analyses

Blood was sampled without stasis through the indwelling antecubital cannula using iced syringes. Plasma glucose and lactate were analysed by standard enzymic methods (Bergmayer, 1974).

The starch analyses were performed at the Dunn Clinical Nutrition Centre, Cambridge using the methods described by Englyst *et al.* (1992) and Englyst & Cummings (1988).

Statistical analyses

Differences between the curves after two identical test meals A1/A2 and B1/B2 were compared by parametric analysis of variance (ANOVA) using a split-plot analysis with time and diet as factors and subjects nested into diet. Thus subjects are regarded as whole-plots and time measurements within subjects as sub-plots. Tests of repeatability were performed according to Bland & Altman (1986). Thus the coefficient of repeatability (CR) = $2 \times$ standard deviation (2 SD) on the differences between meal 1 and meal 2 was calculated for fasting values, 5 h means, 5 h Δ -means (area under the curve (AUC) and Δ -AUC for glucose and lactate), peaks or nadirs and Δ -peaks or nadirs. The CR is defined so that 95% of the difference is expected to be less than 2 SD (British Standards Institution, 1979). To illustrate the use of CR the differences in mean 5 h satiety ratings after meal B1 and B2 are plotted against the mean of the two meals together with the CR (Fig. 1, right panel).

The coefficient of variation (CV) was calculated as:

$$CV = \frac{\sqrt{\frac{(\text{difference})^2}{2 \times \text{number of pairs}}}}{\text{mean}}$$

Difference = result meal 1 – result meal 2.

The CV values were not calculated for the changes from fasting levels (Δ -values).

For illustration, regression analyses were also performed between meals 1 and 2. Total and net AUC for the 5 h measurement periods were calculated separately for each subject as the difference between the integrated area of the response curve and the rectangular area determined by the basal values (trapezoidal method). Statgraphics software version 4.2 (Graphic Software Systems, Inc, Rockville, MD, USA) and the Statistical Analysis Package (SAS Institute, Cary, NC, USA) were used in the statistical calculations.

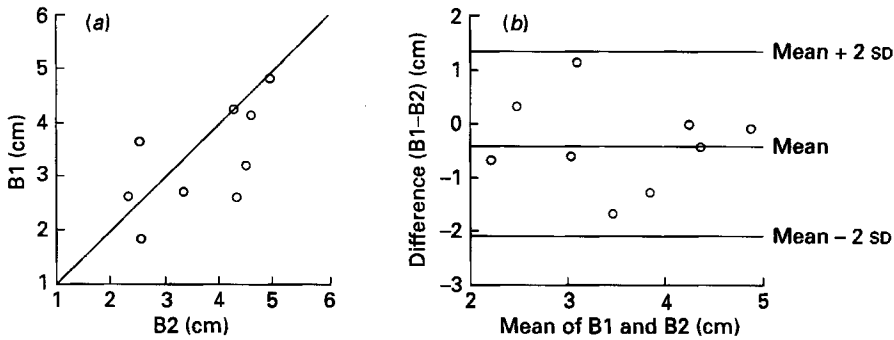


Fig. 1. Mean 5 h satiety scores after meal B taken on two occasions (1 and 2). (a), r^2 0.41, $P = 0.06$ and line of unity; (b), difference in rating for $B1 - B2 \pm 2$ sd (coefficient of repeatability) plotted against mean rating for B1 and B2.

The difference between meal types A and B has been presented previously using data from A2 and B2 (Raben *et al.* 1994c). Data from the ventilated hood measurements will be published separately (Tagliabue *et al.* 1995).

RESULTS

The response curves for the four appetite scores, plasma glucose and lactate concentrations after the test meals are presented in Fig. 2 (A1/A2) and Fig. 3 (B1/B2). The profiles were similar after meals 1 and 2 with no statistically significant differences between the repeated meals as tested by ANOVA.

Repeatability

Fasting values. Mean differences (meal 1 – 2) in fasting appetite scores, and plasma lactate before meals A and B were not significantly different from 0 (Table 1), but fasting plasma glucose concentration was higher before B2 compared with B1 ($P < 0.01$). The CR values were quite large for the appetite scores, ranging between 2.86 and 3.36 cm for meal A and 2.88 and 5.24 cm for meal B and with CV values of 15.6 to 50.5% (Table 1). For fasting glucose, CR values of 0.86 mmol/l (A) and 0.42 mmol/l (B) were found (CV, A: 6.3%; B: 5.3%), while fasting lactate concentrations varied more, showing CR values of 0.42 mmol/l (A) and 1.08 mmol/l (B) (CV, A: 15.9%; B: 41.7%) (Table 1).

Mean appetite scores. Table 2 shows the differences in total and Δ -mean 5 h scores. After meal A the CR values for total mean scores ranged between 1.36 cm (fullness) and 3.80 cm (prospective consumption) and after meal B between 1.62 cm (fullness) and 1.88 cm (prospective consumption). The CV values ranged between 9.6% (hunger B) and 24.9% (satiety A), and the correlation coefficients (r^2) ranged between 0.00 and 0.71.

From Table 2 it can further be seen that the CR values grow larger when using Δ -mean 5 h scores instead of total mean 5 h scores. Thus the CR values ranged between 2.98 cm (Δ -fullness A) and 5.42 cm (Δ -hunger B).

Also for peak/nadir scores, CR values ranging from 3.16 cm (A, hunger) to 5.36 cm (B, satiety) were found (Table 3). Again the CR values increased for all scores when using Δ -values instead of total peak/nadir values. The correlation coefficients between scores after meals 1 and 2 were low for both meals (r^2 0.00–0.30).

When comparing the CR values for all the VAS results, fullness and satiety generally produced lower CR compared with hunger and prospective consumption.

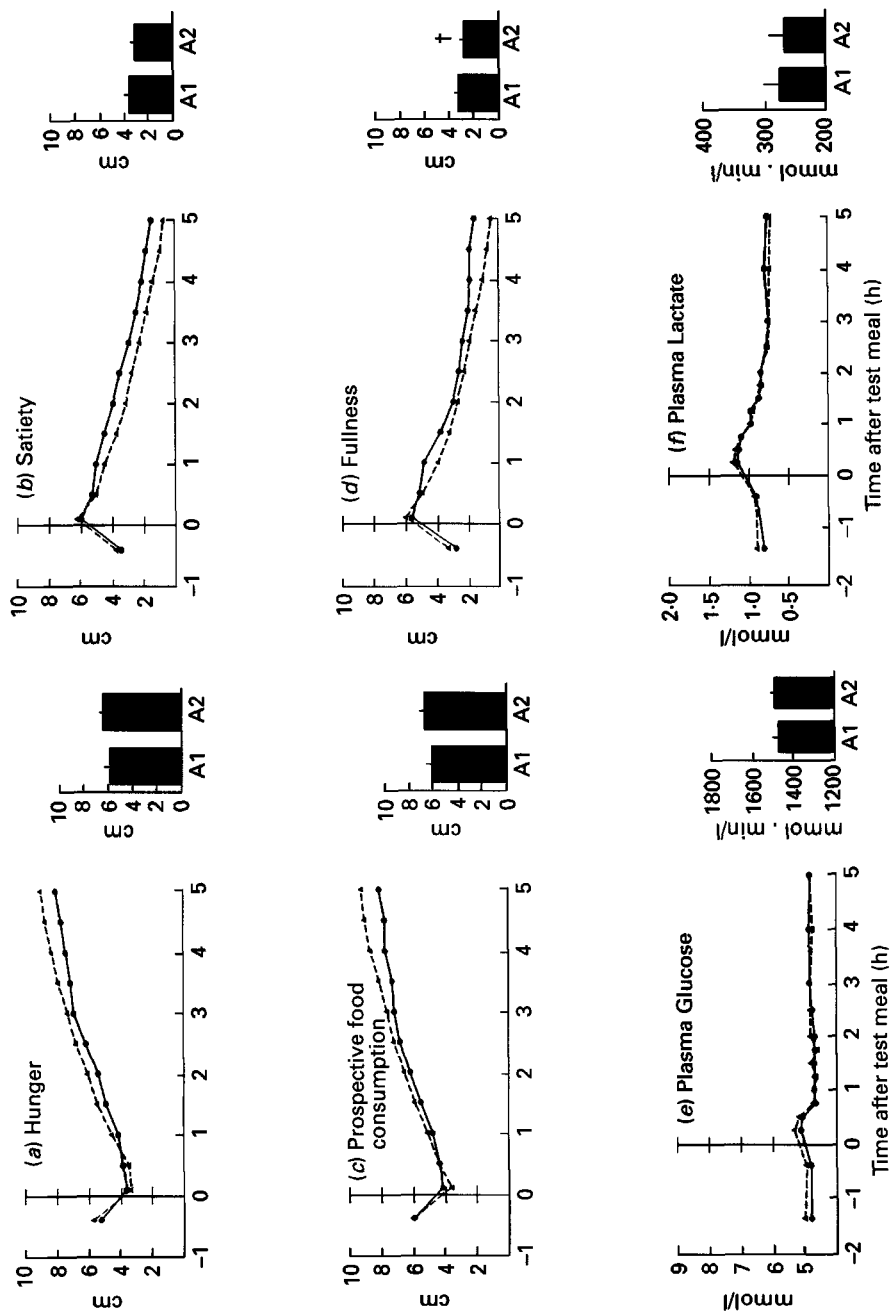


Fig. 2. Subjective appetite scores, plasma glucose and lactate concentrations after meal A taken on two occasions (1, —●— and 2, —▲—) by nine healthy, normal weight, male subjects. Values are means with standard errors indicated by vertical bars. Analysis of variance, all: time effect, $P < 0.0001$; meal effect, not significant. † $P = 0.06$, paired t test. AUC, area under the curve. (a), Hunger; (b), satiety; (c), prospective food consumption; (d), fullness; (e), plasma glucose; (f), plasma lactate. 5 h means are shown alongside (a)–(d), and 5 hr AUC alongside (e)–(f).

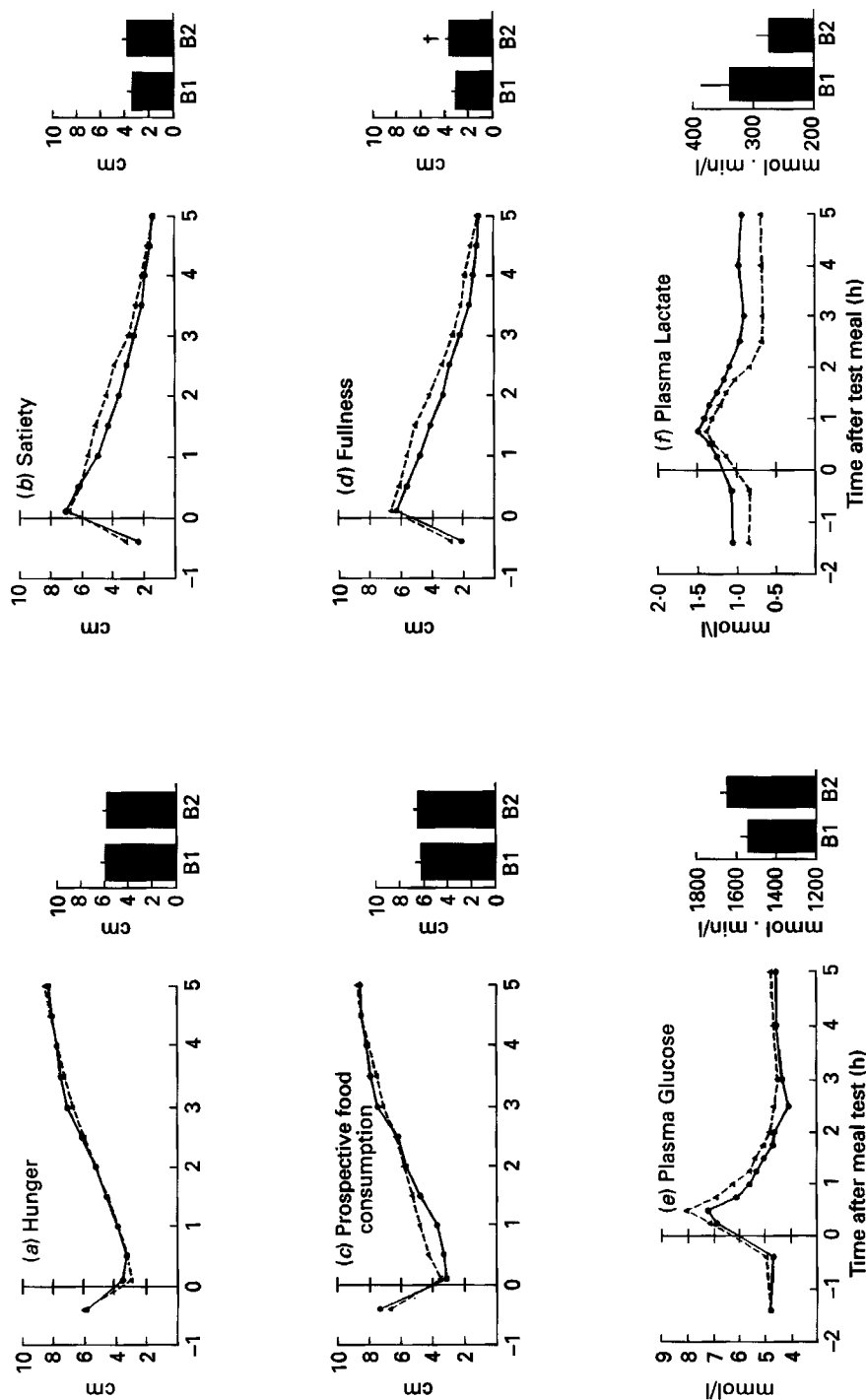


Fig. 3. Subjective appetite scores, plasma glucose and lactate concentrations after meal B taken on two occasions (1, —●— and 2, —▲—) by nine healthy, normal weight, male subjects. Values are means with standard errors indicated by vertical bars. Analysis of variance, all: time effect, $P < 0.0001$; meal effect, not significant. † $P = 0.07$, paired t test. AUC, area under the curve. (a), Hunger; (b), satiety; (c), prospective food consumption; (d), fullness; (e), plasma glucose; (f), plasma lactate. 5 h means are shown alongside (a)–(d), and 5 h AUC alongside (e)–(f).

Table 1. Differences in fasting appetite scores and plasma concentrations of glucose and lactate between repeated tests (1 and 2) of two meals (A and B) containing starch†

	Difference (meal A1 – A2)				Difference (meal B1 – B2)			
	Mean	CR	CV (%)	r ²	Mean	CR	CV (%)	r ²
Hunger (cm)	-0.46	3.66	23.8	0.23	-0.18	5.24	29.7	0.00
Satiety (cm)	-0.32	3.84	36.8	0.28	-0.76	2.88	41.4	0.00
Fullness (cm)	-0.50	2.86	33.3	0.54*	-0.67	3.24	15.6	0.09
Prospective consumption (cm)	-0.03	3.35	19.0	0.06	0.64	3.82	50.5	0.07
Glucose (mmol/l)	-0.14	0.86	6.3	0.04	-0.26**	0.42	5.3	0.14
Lactate (mmol/l)	-0.03	0.42	15.9	0.63*	0.23	1.08	41.7	0.00

CR, coefficient of repeatability = $2 \times \text{SD}$ on mean differences. CV, coefficient of variation.

* $P < 0.05$, ** $P < 0.01$.

† For details of subjects and procedures, see pp. 518–520.

Table 2. Differences in mean and Δ -mean 5 h appetite scores between repeated tests (1 and 2) of two meals (A and B) containing starch†

	Difference (meal A1 – A2)				Difference (meal B1 – B2)			
	Mean	CR	CV (%)	r ²	Mean	CR	CV (%)	r ²
Mean hunger (cm)	-0.59	2.22	14.1	0.14	0.08	1.66	9.6	0.60*
Mean satiety (cm)	0.56	2.12	24.9	0.17	-0.37	1.68	17.6	0.41
Mean prospective consumption (cm)	-0.61	3.80	19.7	0.01	-0.13	1.88	10.2	0.16
Mean fullness (cm)	0.50†	1.36	21.2	0.71**	0.57†	1.62	21.1	0.29
Δ -Mean hunger (cm)	-0.13	4.74		0.00	0.26	5.42		0.01
Δ -Mean satiety (cm)	0.89	4.10		0.04	0.39	3.30		0.18
Δ -Mean prospective consumption (cm)	-0.58	4.40		0.11	-0.78	3.58		0.58*
Δ -Mean fullness (cm)	1.00†	2.98		0.19	0.12	3.56		0.14

CR, coefficient of repeatability = $2 \times \text{SD}$ on mean differences; CV, coefficient of variation.

* $P < 0.05$, ** $P < 0.01$. † $0.05 < P < 0.10$.

‡ For details of subjects and procedures, see pp. 518–520.

Palatability scores. Mean ratings for visual appeal, taste, aftertaste and overall palatability after the four test meals are shown in Fig. 4. Apparently the ratings were very similar after the meals. After meal B, however, the difference in ratings for visual appeal approached significance ($P = 0.053$). The CR values for the four scores were large after both meals, but varied most after A (Table 4). Thus the CR values ranged between 2.38 cm (taste) and 8.70 cm (aftertaste) after A and between 3.58 cm (visual appeal) and 4.86 cm (taste) after B. The mean differences (meal 1–2) were negative in seven out of the total eight scores.

Glucose and lactate. The profiles of total plasma glucose and lactate concentrations are shown in Figs. 1 and 2. No significant meal effects were found, although slightly different levels were found after meal B (Fig. 2). The CR values for the total 5 h AUC averaged 176.0 mmol. min/l (A) and 208.2 mmol. min/l (B), i.e. small ranges compared with the total areas (CV A: 4.0%, B: 6.2%) (Table 5). However, the CR values for the Δ -AUC were much larger compared with those for the mean AUC.

Table 3. Differences in total and Δ -peak/nadir appetite scores between repeated tests (1 and 2) of two meals (A and B) containing starch†

	Difference (meal A1 – A2)				Difference (meal B1 – B2)			
	Mean	CR	CV (%)	r^2	Mean	CR	CV (%)	r^2
Nadir hunger (cm)	0.21	3.16	36.3	0.06	-0.16	3.48	42.4	0.21
Peak satiety (cm)	0.13	3.58	19.0	0.10	0.30	5.36	26.3	0.03
Nadir prospective consumption (cm)	0.24	4.06	37.9	0.02	-0.52	4.76	51.9	0.01
Peak fullness (cm)	0.03	3.82	21.2	0.10	-0.33	4.00	21.2	0.01
Δ -Nadir hunger (cm)	0.67	5.56		0.04	0.02	5.48		0.07
Δ -Peak satiety (cm)	0.47	4.80		0.07	1.06	6.44		0.03
Δ -Nadir prospective consumption (cm)	0.27	5.02		0.01	-1.18	5.32		0.09
Δ -Peak fullness (cm)	0.53	3.88		0.30	0.34	4.90		0.14

CR, coefficient of repeatability = $2 \times \text{SD}$ on mean differences; CV, coefficient of variation.

† For details of subjects and procedures, see pp. 518–520.

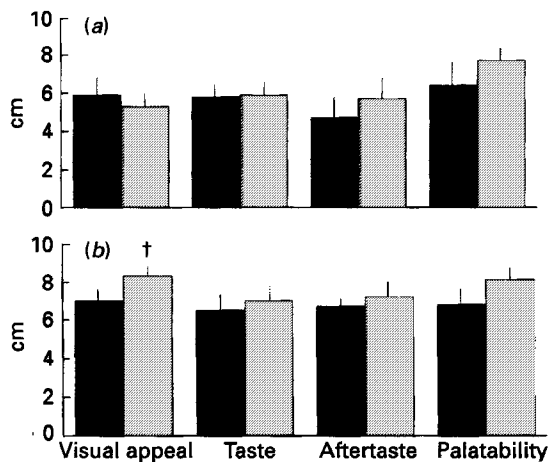


Fig. 4. Subjective evaluation of test meals A (a) and B (b) on two occasions, 1 (■) and 2 (▨). Values are means with their standard errors indicated by vertical bars (n 9). † $P = 0.053$, paired t test.

For peak glucose, CR values of 0.76 mmol/l (A) and 1.46 mmol/l (B) were found. The CR increased when comparing Δ -peak values for A (0.92 mmol/l), but remained the same for B (1.46 mmol/l).

For lactate the CR values for 5 h AUC averaged 134.8 mmol.min/l (A) and 293.6 mmol.min/l (B), (CV A: 16.9%, B: 35.6%) (Table 5). For Δ -AUC, the CR values increased compared with the mean Δ -AUC.

Correlation analyses: VAS scores, glucose and lactate

Correlation analyses on fasting values before meal A showed significant correlations between differences in fasting plasma glucose and fullness (r 0.76, $P = 0.02$) and satiety (r 0.65, $P = 0.06$) and between lactate and prospective consumption (r -0.65, $P = 0.06$). No

Table 4. Differences in subjective scores between repeated tests (1 and 2) of two meals (A and B) containing starch†

	Difference (meal A1 – A2)				Difference (meal B1 – B2)			
	Mean	CR	CV (%)	r ²	Mean	CR	CV (%)	r ²
Visual appeal (cm)	0.57	3.06	19.6	0.77**	-1.36†	3.58	20.0	0.29
Taste (cm)	-0.12	2.38	13.6	0.80**	-0.46	4.86	24.5	0.38
Aftertaste (cm)	-0.94	8.70	57.2	0.04	-0.44	4.48	21.9	0.16
Palatability (cm)	-1.26	6.28	32.2	0.00	-1.31	4.74	24.6	0.22

CR, coefficient of repeatability = $2 \times \text{SD}$ on mean differences; CV, coefficient of variation.

** $P < 0.01$, † $0.05 < P < 0.10$.

† For details of subjects and procedures, see pp. 518–520.

Table 5. Differences in peaks and areas under the curves (AUC) for plasma glucose and lactate after two test meals A and B on two occasions, 1 and 2‡

	Difference (meal A1 – A2)				Difference (meal B1 – B2)			
	Mean	CR	CV (%)	r ²	Mean	CR	CV (%)	r ²
Glucose								
AUC (mmol.min/l)	-14.7	176.0	4.0	0.19	-97.8*	208.2	6.2	0.33
Δ -AUC (mmol.min/l)	28.7	163.0		0.02	-19.0	190.2		0.19
Peak (mmol/l)	-0.25†	0.76	5.8	0.17	-0.57*	1.46	8.1	0.57*
Δ -Peak (mmol/l)	-0.10	0.92		0.06	-0.32	1.46		0.58*
Lactate								
AUC (mmol.min/l)	10.3	134.8	16.9	0.48*	66.1	293.6	35.6	0.05
Δ -AUC (mmol.min/l)	18.9	105.6		0.36	-5.0	87.2		0.46*
Peak (mmol/l)	-0.03	0.60	16.5	0.27	0.04	0.80	18.1	0.13
Δ -Peak (mmol/l)	-0.00	0.26		0.58*	-0.19	0.74		0.01

CR, coefficient of repeatability = $2 \times \text{SD}$ on mean differences; CV, coefficient of variation.

* $P < 0.05$, † $0.05 < P < 0.10$.

‡ For details of subjects and procedures, see pp. 518–520.

correlations were found before B. When correlating the differences after meals 1 and 2 a positive correlation was found between Δ -peak glucose and Δ -mean fullness for A (r 0.72, $P = 0.03$). For B, positive correlations were found between differences in Δ -AUC lactate and Δ -nadir hunger (r 0.71, $P = 0.04$), Δ -mean hunger (r 0.70, $P = 0.04$) and Δ -mean prospective consumption (r 0.65, $P = 0.06$). Differences in AUC glucose and nadir prospective consumption for B were negatively correlated ($r = -0.72$, $P = 0.05$), as were Δ -peak lactate and Δ -mean satiety ($r = -0.80$, $P = 0.01$).

A number of correlations were found between the differences in subjective evaluation of the test meal, satiety ratings and glucose/lactate responses. Thus for meal A, differences in overall palatability were correlated with mean hunger (r 0.67, $P = 0.05$), prospective consumption (r 0.69, $P = 0.04$), Δ -mean fullness ($r = -0.68$, $P = 0.04$) and with AUC glucose (r 0.69, $P = 0.04$), and differences in taste ratings with Δ -mean fullness ($r = -0.68$, $P = 0.04$). Aftertaste was correlated with AUC lactate ($r = -0.67$, $P = 0.05$), Δ -AUC lactate ($r = -0.78$, $P = 0.01$) and peak glucose (r 0.69, $P = 0.04$). From a stepwise selection analysis for Δ -mean fullness including Δ -glucose, Δ -peak glucose, taste and palatability, Δ -peak glucose was the only factor remaining significant (r 0.68, $P = 0.03$). After meal B, differences in

aftertaste ratings were correlated with mean satiety ($r = 0.67$, $P = 0.049$), peak satiety ($r = 0.72$, $P = 0.03$), peak fullness ($r = 0.90$, $P = 0.001$) and Δ -peak fullness ($r = 0.66$, $P = 0.052$).

DISCUSSION

The present study shows that, although similar profile curves were found postprandially, subjective appetite scores are not easily reproduced when tested after identical meals in the same subjects on different days under standardized conditions. This was true whether expressed as fasting, mean or peak/nadir ratings. The use of incremental values, which eliminates differences due to a possible variation in fasting ratings, did not change these findings, but instead increased the CR. Thus the lowest CR values were found when mean 5 h values were used compared with the other measures of appetite.

We still find, however, that the CR values for mean 5 h values were high on the scale from 0 to 10 cm. To illustrate this the range for mean difference on the score with the lowest CR can be calculated. Mean 5 h hunger ratings for B1 and B2 averaged 5.83 and 5.75 cm respectively, i.e. a small difference between B1 and B2 of 0.08 cm. However, the CR was 1.66 cm, i.e. a total range of 3.32 cm (Fig. 1). At first glance this does not seem acceptable on a scale from 0 to 10 cm. We are aware, however, that the limited number of subjects included in the present study ($n = 9$) may be part of the explanation for the large CR values. Even for this group of subjects, however, we expected the repeatability of the different scores to be better. Moreover, a sample size of nine subjects is not rare in studies on appetite sensations.

The glucostatic concept has recently been reappraised as an important physiological factor influencing appetite. According to this concept the control of appetite is to a great extent influenced by the amount of available carbohydrate in the diet (Flatt, 1987; Blundell & Burley, 1990; van Amelsvoort *et al.* 1990) and the consumption of a certain amount of carbohydrate (or protein) in a meal may be necessary to achieve satiation (Tremblay *et al.* 1991; Astrup & Raben, 1992). In contrast, a depletion of the glycogen stores or a lowering of blood glucose concentration will elicit neural signals to the satiety centre in the brain giving rise to food consumption (Forbes, 1992). In the present study carbohydrate was the only energy source in the test meals. This may therefore be expected to increase the possibility of finding any correlations between differences in appetite scores and plasma glucose concentration, as the interference from fat and protein is then eliminated. We therefore correlated the differences in scores after meals 1 and 2 with the differences in glucose and lactate responses. Some correlations were in fact found, but not to a degree able to explain the overall large intra-subject variation in appetite scores.

The lack of consistent correlations between differences in appetite scores and plasma glucose concentrations in the present study corresponds to the findings from our previous meal test studies, where we were not able to find correlations between appetite ratings and plasma glucose concentrations (Raben *et al.* 1994*a, b, c*). The lack of correlations in the present study therefore should probably not be interpreted as a failure on the part of the VAS scores, but rather as evidence that hunger and satiety are not solely dependent on glucose metabolism.

Appetite and satiety sensations may be influenced by a variety of different factors other than plasma glucose, one of these being the palatability of the test meal (Blundell, 1979; Forbes, 1992). In particular, appetite may be enhanced if food is perceived as pleasant (Blundell, 1979). The test meals used in the present study were not representative of an everyday diet, but represented an experimental situation in which non-familiar foods were tested. This may have been reflected in the palatability of meal A being rated as medium, while meal B was given an even less positive evaluation (Table 3). It is possible that the fact

that the subjects in seven out of eight cases rated the palatability of test meals worse on the second occasion compared with the first occasion can explain part of the low repeatability of the appetite scores. Significant correlations were found between differences in overall palatability and hunger, prospective consumption and fullness ratings for A, showing that hunger and food consumption decreased and fullness increased when palatability decreased. After meal B differences in aftertaste were negatively correlated with peak satiety and fullness. Together these findings suggest that palatability may in fact have had an impact on the different appetite sensations.

Although the subjects did not know beforehand which meal they were given on the test days, the meals were not covertly manipulated. Thus the prior experience of the test meals may have created an order-effect in their responses, resulting in a conditioned satiation or hunger.

A true biological day-to-day variation in subjective hunger sensations may also explain the low degree of reproducibility of the scores. An impression of such a variability may be gained from the fasting scores. The CR values here were in fact large and of a similar size compared with the 5 h mean or peak/nadir appetite scores. This may therefore support the hypothesis of true day-to-day variation in appetite sensations. The variation in the VAS ratings due to methodological problems cannot, however, be distinguished from the variation caused by biological determinants of appetite and satiety. To help circumvent the problem of variation, randomization of the order of test meals in studies assessing subjective appetite is recommended.

In conclusion, large reproducibility coefficients for subjective appetite ratings were found after identical meals, when tested twice in the same subjects on different days under standardized pre-test conditions. Differences in palatability ratings or the prior experience of the meal resulting in a conditioned satiation or hunger may partly explain the large intra-subject variation in appetite ratings. However, it is likely that both a substantial methodological day-to-day variability and a biological day-to-day variation in subjective hunger and appetite sensations may have been involved. Further studies including more subjects were needed to assess the reproducibility of subjective appetite scores.

The authors gratefully thank Lene Mellekjær Jensen, the laboratory technicians Bente Knap, Inge Timmermann, John Lind, Claudia Trentani and Inger-Lise Grønfeldt and her kitchen staff for expert technical assistance. The study was supported by grants from the Danish Research and Development programme for Food Technology 1990–1994 and the Danish Medical Research Council, grant no. 12-9537-3.

REFERENCES

- Astrup, A. & Raben, A. (1992). Obesity: an inherited metabolic deficiency in the control of macronutrient balance? *European Journal of Clinical Nutrition* **46**, 611–620.
- Bergmayer, H. V. (editor) (1974). *Methods of Enzymatic Analysis*. New York: Academic Press.
- Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**, 307–310.
- Blundell, J. E. (1979). Hunger, appetite and satiety – constructs in search of identities. In *Nutrition and Lifestyles*, pp. 21–42 [M. Turner, editor]. London: Applied Science Publishers.
- Blundell, J. E. (1990). Appetite disturbance and the problems of overweight. *Drugs* **39**, Suppl. 3, 1–19.
- Blundell, J. E. & Burley, V. J. (1990). Evaluation of the satiating power of dietary fat in man. In *Progress in Obesity Research 1990*, pp. 453–457 [Y. Oomura, S. Tarui, S. Inoue and T. Shimazu, editors]. London: John Libbey.
- Blundell, J. E. & Hill, J. J. (1988). On the mechanism of action of dexfenfluramin: effect on alliesthesia and appetite motivation in lean and obese subjects. *Clinical Neuropharmacology* **11**, S121–S134.
- Bolton, R. P., Heaton, K. W. & Burroughs, L. F. (1981). The role of dietary fiber in satiety, glucose, and insulin: studies with fruit and fruit juice. *American Journal of Clinical Nutrition* **34**, 211–217.

- British Standards Institution (1979). *Precision of Test Methods. I. Guide for the Determination and Reproducibility for a Standard Test Method* (BS 5497, part 1). London: BSI.
- Burley, V. J., Paul, A. W. & Blundell, J. E. (1993). Sustained post-ingestive action of dietary fibre: effects of a sugar-beet-fibre-supplemented breakfast on satiety. *Journal of Human Nutrition and Dietetics* **6**, 43–50.
- Costill, D. L. (1988). Carbohydrates for exercise: dietary demands for optimal performance. *International Journal of Sports Medicine* **9**, 1–18.
- Deurenberg, P., van der Kooy, K., Leenen, R., Westrate, J. A. & Seidell, J. C. (1991). Sex and age specific prediction formulas for estimating body composition from bioelectrical impedance: a cross-validation study. *International Journal of Obesity* **15**, 17–25.
- Dixon, J. S. & Bird, H. A. (1981). Reproducibility along a 10 cm vertical visual analogue scale. *Annals of the Rheumatic Diseases* **40**, 87–89.
- Englyst, H. N. & Cummings, J. H. (1988). An improved method for the measurement of dietary fibre as non-starch polysaccharides in plant foods. *Journal of the Association of Official Analytical Chemists* **71**, 808–814.
- Englyst, H. N., Kingman, S. M. & Cummings, J. H. (1992). Classification and measurement of nutritionally important starch fractions. *European Journal of Clinical Nutrition* **46**, Suppl. 2, S33–S50.
- Flatt, J. P. (1987). Dietary fat, carbohydrate balance, and weight maintenance. *American Journal of Clinical Nutrition* **45**, 296–306.
- Food and Agriculture Organization/World Health Organization/United Nations University (1985). *Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. Technical Report Series no 74*. Geneva: World Health Organization.
- Forbes, J. M. (1992). Metabolic aspects of satiety. *Proceedings of the Nutrition Society* **51**, 13–19.
- Haber, G. B., Heaton, K. W., Murphy, D. & Burroughs, L. F. (1977). Depletion and disruption of dietary fibre: effects on satiety, plasma-glucose, and serum-insulin. *Lancet* **ii**, 679–682.
- Hill, A. J., Magson, L. D. & Blundell, J. E. (1984). Hunger and palatability: tracking ratings of subjective experience before, during and after the consumption of preferred and less preferred food. *Appetite* **5**, 361–371.
- Joyce, C. R., Zutshi, D. W., Hrubes, V. & Mason, R. M. (1975). Comparison of fixed interval and visual analogue scales for rating chronic pain. *European Journal of Clinical Pharmacology* **8**, 415–420.
- Lahuerta, D. R., Campbell, J. & Dewqshi, F. (1983). Muscular ischemic pain. Effect of 'training' after repeated performance of an experimental test. *Revista de Medicina de la Universidad de Navarra* **27**, 41–45.
- Lappalainen, R., Mennen, L., van Weert, L. & Mykkänen, H. (1993). Drinking water with a meal: a simple method of coping with feelings of hunger, satiety and desire to eat. *European Journal of Clinical Nutrition* **47**, 815–819.
- Leathwood, P. & Pollet, P. (1988). Effects of slow release carbohydrates in the form of bean flakes on the evolution of hunger and satiety in man. *Appetite* **10**, 1–11.
- Loiseau, A., Dubreuil, C. & Pujet, J. C. (1990). A visual analog scale of exercise dyspnoea. *Revue des Maladies Respiratoires* **7**, 39–44.
- Mathers, J. C. (1991). Digestion of non-starch polysaccharides by non-ruminant omnivores. *Proceedings of the Nutrition Society* **50**, 161–172.
- Mattes, R. (1990). Hunger ratings are not a valid proxy measure of reported food intake in humans. *Appetite* **15**, 103–113.
- Ohnhaus, E. E. & Adler, R. (1975). Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain* **1**, 379–384.
- Pi-Sunyer, X., Kissileff, H. R., Thornton, J. & Smith, G. P. (1982). C-terminal octapeptide of cholecystokinin decreases food intake in obese men. *Physiology and Behavior* **29**, 627–630.
- Raben, A., Andersen, H. B., Christensen, N. J., Madsen, J., Holst, J. J. & Astrup, A. (1994a). Evidence for an abnormal postprandial response to a high-fat meal in women predisposed to obesity. *American Journal of Physiology* **267**, 549–559.
- Raben, A., Christensen, N. J., Madsen, J., Holst, J. J. & Astrup, A. (1994b). Decreased post-prandial thermogenesis and fat oxidation, but increased fullness after a high-fiber meal compared to a low-fiber meal. *American Journal of Clinical Nutrition* **59**, 1386–1394.
- Raben, A., Tagliabue, A., Christensen, N. J., Madsen, J., Holst, J. J. & Astrup, A. (1994c). Resistant starch: the effect on postprandial glycemia, hormonal response and satiety. *American Journal of Clinical Nutrition* **60**, 544–551.
- Rogers, P. J., Carlyle, J., Hill, A. J. & Blundell, J. E. (1988). Uncoupling sweet taste and calories: comparison of the effects of glucose and three intense sweeteners on hunger and food intake. *Physiology and Behavior* **43**, 547–552.
- Rolls, B. J., Hetherington, M. & Laster, L. J. (1988). Comparison of the effects of aspartame and sucrose on appetite and food intake. *Appetite* **11**, 62–67.
- Silverstone, T. & Goodall, E. (1986). Measurement of hunger and food intake. In *Disorders of Eating Behaviour, a Psychoneuroendocrine Approach*, pp. 129–134 [E. Ferrari, editor]. Oxford: Pergamon Press.
- Silverstone, T. & Stunkard, A. J. (1968). The anorectic effect of dexamphetamine sulphate. *British Journal of Pharmacology and Chemotherapy* **33**, 513–522.
- Spiegel, T. A., Stunkard, A. J., Shrager, E. E., O'Brian, C. P., Morrison, M. F. & Stellar, E. (1987). Effect of naltrexone on food intake, hunger and satiety in obese men. *Physiology and Behavior* **40**, 135–141.

- Tagliabuc, A., Raben, A., Heijnen, M.-L., Deurenberg, P., Pasquali, E. & Astrup, A. (1995). The effect of resistant starch on energy expenditure and substrate oxidation. *American Journal of Clinical Nutrition* (In the Press).
- Tremblay, A., Lavallée, N., Almeras, N., Allard, L., Després, J.-P. & Bouchard, C. (1991). Nutritional determinants of the increase in energy intake associated with a high-fat diet. *American Journal of Clinical Nutrition* **53**, 1134–1137.
- van Amelsvoort, J. M. M., van Stratum, P., Kraal, J. F., Lussenburg, R. N. & Dubbelman, G. P. (1990). Minor difference in postprandial responses of men between starch and sugar when replacing fat in a normal meal. *British Journal of Nutrition* **63**, 37–51.