Reproducibility of short-term food intake and subjective appetite scores after a glucose preload, ventilation threshold, and body composition in boys

Nick Bellissimo, Scott G. Thomas, Paul B. Pencharz, Robert C. Goode, and G. Harvey Anderson

Abstract: The objective of these studies was to assess the reproducibility of (i) short-term food intake (FI) and subjective appetite following a glucose preload, (ii) ventilation threshold (VT) and subjective appetite after short-duration exercise, and (iii) body composition assessed by bioelectrical impedance (BIA). On two separate weekend mornings, boys (n = 11; aged 9–14 years) received drinks containing 50 g glucose made up to 250 mL with water 2 h after a standardized breakfast. FI from a pizza meal was measured 30 min later. Subjective appetite was measured before and after the glucose loads and the VT measures. VTs were measured on 2 weekday evenings, 1 week apart. BIA was measured during the FI assessment sessions. Short-term FI after the glucose preload was highly reproducible. Mean energy intake was 925 ± 139 kcal on the first day and 988 ± 147 kcal on the second day (coefficient of repeatability (CR) = 259 kcal; intra-class correlation coefficient (ICC) = 0.96). Moderate reproducibility of the average appetite score was found at 30 min (CR = 24 mm; ICC = 0.82). Subjective appetite was increased similarly by short-duration exercise on both days (CR = 19 mm). Absolute VT was more highly reproducible (CR = 359 mL O₂·min⁻¹, ICC = 0.85) than VT expressed on the basis of body weight (CR = 8.0 mL O₂·kg⁻¹·min⁻¹, ICC = 0.59). Fat mass (FM) estimated from BIA was highly reproducible (CR = 2.7 kg, ICC = 0.95), but underestimated FM compared with skinfolds. In conclusion, FI and subjective appetite scores in response to glucose preloads, VT, subjective appetite after short-duration exercise, and estimates of FM from BIA are reproducible in boys.

Key words: bioelectrical impedance analysis, children, food intake, reproducibility, subjective appetite, ventilation threshold.

Résumé : Le but de ces études est d’évaluer la reproductibilité (i) d’un apport alimentaire (FI) récent et de la sensation de faim après un apport de glucose, (ii) du seuil ventilatoire (VT) et de la sensation de faim après un effort de courte durée et (iii) de la composition corporelle mesurée par l’analyse de l’impédance bioélectrique (BIA). Durant la matinée de deux fins de semaine, onze garçons âgés de 9 à 14 ans absorbent 250 mL d’une boisson contenant de l’eau et 50 g de glucose 2 h après un petit déjeuner standard. Trente minutes après avoir mangé une pointe de pizza, on évalue la FI. De plus, on évalue la sensation de faim avant et après l’apport de glucose et les mesures de VT. Et on détermine les VT au cours de deux soirées sur semaine, et ce, à une semaine d’intervalles. L’impédance bioélectrique (BIA) est analysée durant les séances d’évaluation du FI. La sensation de faim après un apport récent de glucose est très reproductible. Le premier jour, l’apport énergétique moyen est de 925 ± 139 kcal et le deuxième jour, de 988 ± 147 kcal : (coefficient de reproductibilité, CR = 259 kcal; coefficient de corrélation intraclass, ICC = 0.96). La valeur du coefficient de reproductibilité de la sensation moyenne de faim mesuré 30 min après avoir mangé est modérée (CR = 24 mm; ICC = 0.82). La sensation de faim augmente de façon comparable après la séance d’exercice physique réalisée en deux jours distincts (CR = 19 mm). La valeur absolue du VT est plus reproductible (CR = 359 mL O₂·min⁻¹, ICC = 0.85) que cette même valeur exprimée en fonction de la masse corporelle (CR = 8.0 mL O₂·kg⁻¹·min⁻¹, ICC = 0.59). La masse adipeuse (FM) estimée d’après la BIA est très reproductible (CR = 2.7 kg, ICC = 0.95), mais donne une valeur inférieure à celle obtenue au moyen de la technique des plis cutanés. En conclusion, l’apport alimentaire et la sensation de faim évalués après un apport de glucose, le VT, la sensation de faim après un exercice physique de courte durée et les valeurs de la masse adipeuse obtenues par l’analyse de l’impédance bioélectrique sont des données reproductibles chez les garçons.
Introduction

With nearly one in three Canadian children overweight (OW) or obese (OB) (Tremblay et al. 2002), an understanding of the significance of interactions between food intake (FI) regulation, subjective appetite, exercise and fitness, and body composition is required for the development of guidelines for promoting healthier body weights in children. However, the reproducibility of test methods aimed at measuring FI regulation, subjective appetite, physical fitness, and body composition has received little investigation in children. The correlation coefficient between two identical trials is often reported to assess reproducibility; however, the use of correlations is misleading and not sufficient to describe reproducibility of a method. This requires calculation of the coefficient of repeatability (CR = 2 × standard deviation (SD) on the difference between the repeated measurements) (Bland and Altman 1986) and the intra-class correlation coefficient (ICC) (Streiner and Norman 1989).

The role of FI regulation, subjective appetite, exercise and fitness, and body composition as determinants of body weight are of current interest. To assess short-term FI control, a preload design is often used in which treatments of fixed energy content (caloric preload) are provided and FI is measured at a later ad libitum meal (Anderson and Woodend 2003; Bellissimo et al. 2007a; Blundell et al. 1994). The reproducibility of the effect of consuming a preload on FI at a later meal has not been reported in children. FI in adults at an ad libitum lunch-time meal consumed after a fixed-sized breakfast on two separate days was almost identical, implying that reproducibility was high (Flint et al. 2000). However, because the CR was not reported for FI it is unknown whether there were large within-subject variations in energy intake at the test meals.

Subjective appetite sensations assessed by visual analogue scales (VAS) often provide important additional information when evaluating the impact of caloric preloads on satiety. Reproducibility in adult subjective appetite scores measured consecutively on the same day show high reproducibility (Stratton et al. 1998), but are lower when measured on different days (Flint et al. 2000; Raben et al. 1995). However, there are no reports of the reproducibility of VAS in children.

Despite its common use, there are few reports of the reproducibility of the VT in children in response to identical exercise protocols on separate occasions (Mahon and Marsh 1992; Weymans and Reybrouck 1989). The VT is a commonly used marker of aerobic fitness in children (Cooper et al. 1984; Paterson et al. 1987; Rowland and Green 1988) because it is noninvasive, does not require a maximal effort, and is a sensitive indicator of physical performance at a heart rate of 170 beats min\(^{-1}\) (Reybrouck et al. 1986). The VT can be also determined without difficulty by multiple observers in adults (Caiozzo et al. 1982) and children (Hebestreit et al. 2000; Ohuchi et al. 1996).

Bioelectrical impedance analysis (BIA) is a commonly used field method for estimating fat mass (FM) and has shown to be reproducible in children, often replacing skinfold measures. The Horlick equation (Horlick et al. 2002) is a newly developed prediction equation for estimating body composition from BIA that has been cross-validated against deuterium dilution and dual-energy X-ray absorptiometry in a large sample of children and adolescents (4–18 years). The reproducibility of estimates of FM based on the Horlick equation has not been reported. There is some evidence, however, that FM from BIA has poorer precision than skinfolds in estimating FM (Goran et al. 1996).

Therefore, the objective of these studies was to determine the reproducibility of FI and subjective appetite sensations following exercise of short-duration, and FM from BIA, as well as agreement between estimates of FM from BIA and skinfolds in 9- to 14-year-old boys.

Materials and methods

Subjects

Eleven boys aged 9–14 years old (mean ± SEM; 10.8 ± 0.5 years), with a mean body mass index (BMI) of 19.7 ± 0.8 kg m\(^{-2}\) (BMI percentile range was between the 8th and 97th) participated in the study. In Canada, the US Centers for Disease Control (CDC) and Prevention 2000 growth charts are recommended for monitoring growth in children 2 years of age and older (Dietitians of Canada et al. 2004). However, the terminology used in Canada differs from that of the CDC. Normal weight is defined as a BMI that is between the 5th and 85th percentile for age and gender. Overweight is defined as a BMI between the 85th and 95th percentile, whereas obesity is greater than the 95th percentile. The American terminology was chosen to avoid labeling a child as OB; therefore, CDC charts refer to children between the 85th and 95th percentile as “at risk of OW” and above the 95th percentile as “OW”. Subjects were recruited from the University of Toronto Schools via a recruitment letter sent home to parents and by word of mouth. The Human Subjects Review Committee of the Ethics Review Office at the University of Toronto approved this study.

To be included in the study, the boys had to have been born at full-term and of normal weight at birth (confirmed with the parent). Individuals dieting, taking medication, and with any significant learning, behavioral, or emotional difficulties (as reported by the parents but not verified by clinical testing) were excluded from the research as previously reported (Bellissimo et al. 2008; Bellissimo et al. 2007a; Bellissimo et al. 2007b). However, none of the boys appeared to have any significant learning, behavioral, or emotional problems that limited their participation in the study.
Food intake and subjective appetite

A within-subjects repeated measures design was used to examine children’s subjective appetite and FI at a pizza meal 30 min after a glucose preload. The experimental protocol was repeated on two separate weekend mornings, 1 week apart. On each of the two weekend mornings the boys were given equally sweetened preloads of glucose (50 g) made up to 250 mL with water 2 h after a standardized breakfast of milk, cereal, and orange juice. The boys completed VAS measures of subjective appetite prior to the preloads and 15, 30, and 60 min later. They were given a pizza lunch at 30 min and allowed to eat until they were comfortably full.

The preload design used herein is described in greater detail elsewhere (Bellissimo et al. 2008; Bellissimo et al. 2007a). Briefly, the boys arrived at the Department of Nutritional Sciences at 10h00 or 11h00, 2 h after consuming a standardized breakfast of skim milk (250 mL, 381 kJ), Honey Nut Cheerios® (26 g, 430 kJ, donated by General Mills, Inc., Mississauga, Ont.) and Tropicana Orange Juice® (236 mL, 460 kJ) at home (8h00 or 9h00). Upon arrival, subjects completed baseline (0 min) VAS questionnaires measuring their motivation to eat and the VAS were also administered at 15, 30, and 60 min after the preloads.

The glucose preload contained 50 g of glucose (54.6 g of glucose monohydrate; Grain Process Enterprises, Toronto, Ont.) made up to 250 mL with water. Sweetened, cherry-flavored crystals (Sugar Free Kool-Aid, Kraft Canada Inc., Don Mills, Ont.) were also added to improve the flavor and palatability. The preloads were served chilled in an opaque cup, and the boys were instructed to consume the drink within 5 min on both visits. Thirty minutes after the boys consumed the glucose preload they were escorted into the taste panel room and individually seated in their own cubicle, free of external cues, and served a pizza lunch. Pizza was served with a 500 mL bottle of spring water (Danone Crystal Springs, Québec City, Que.). The boys were informed that additional hot trays of pizza would be provided at regular intervals and advised to eat until they were “comfortably full”.

The motivation to eat VAS questionnaire, used to assess appetite, consisted of 4 questions or scales: (i) how strong is your desire to eat? (“very weak” to “very strong”), (ii) how hungry do you feel? (“not hungry at all” to “as hungry as I’ve ever felt”), (iii) how full do you feel? (“not full at all” to “very full”), and (iv) how much food do you think you can eat? (prospective food consumption, PFC; “nothing at all” to “a large amount”) (Bellissimo et al. 2008; Bellissimo et al. 2007a). Each VAS consisted of a 100 mm line anchored at the beginning and end by opposing statements. The subjects marked an “X” on the line to indicate their feelings at the given moment. Scores were determined by measuring the distance (in mm) from the left starting point of the line to the intersection of the “X”.

The sweetness of the glucose preloads was also measured using VAS. The question “How sweet have you found the drinks?” was anchored by “not at all sweet” and “very sweet”.

The ad libitum lunch consisted of pizza. The boys rated the pizza according to their preference before the sessions. Participants were served 2 pizzas of their first choice and 1 pizza of their second choice on each tray. Only when the children indicated a dislike for a particular variety of pizza were they served 3 identical pizzas on each tray. Boys were served the same variety of pizza during each treatment session.

Two varieties of Deep ’n’ Delicious® 5” diameter pizza were fed (averaging 210 kcal per pizza); pepperoni and three-cheese pizzas donated by McCain Foods (McCain Canada Ltd., Florenceville, Ont.). Pepperoni pizza (102 g) contained 9.7 g of protein, 7 g of fat, and 26 g of carbohydrates for a total energy content of 206 kcal. Each three-cheese pizza (96 g) contained 11 g of protein, 7.3 g of fat and 26 g of carbohydrate for a total energy content of 213 kcal. The cooked pizzas were weighed and cut into 4 equal pieces before serving, and the amount left after the meal was subtracted from the initial weight to provide a measure of FI, as previously reported (Bellissimo et al. 2008; Bellissimo et al. 2007a; Bellissimo et al. 2007b).

Physical fitness

Physical fitness of the children was measured in the Human Physiology and Performance Laboratory (Athletic Centre, University of Toronto) on 2 weekday evenings (between 16h00 and 18h00 and at least 3 h after their last meal), 1 week apart, by measuring the VT during a continuous, progressive, pseudo-ramp treadmill protocol (Porszasz et al. 2003; Thomas et al. 2003). Subjective appetite sensations were assessed before and after the fitness test for each subject.

The VT is defined as the exercise load at which ventilation begins to increase exponentially for a given increase in VO₂ and typically occurs at 45%–65% of VO₂ max (Fleg et al. 2000). A walking protocol on a treadmill (Quinton Model 18-60-I) was employed to assess VT in children, as previously described (Bellissimo et al. 2007b). Ventilatory gases were collected from the mouth through a two-way non-rebreathing valve and gas exchange was assessed using a metabolic cart (MOXUS, AEI Tech). A turbine was used to measure inspired ventilation and a sample of mixed-expired gas was passed on to a rapidly responding gas analyzer. Carbon dioxide (CO₂) was analyzed via photo acoustic infrared spectroscopy and oxygen (O₂) via an electrochemical cell. Calibration of CO₂ and O₂ readings was done before and after each fitness test using gases of known concentration.

The treadmill walking protocol was initially set at 1% grade, and a walking speed of 0.89 m·s⁻¹ for the first 3 min. Every minute thereafter, the increase in incline (0.5%–2.5%) and speed (0.09–0.22 m·s⁻¹) depended on the protocol (initial speed and incline after the third minute: low, incline = 7.5% and speed = 0.98 m·s⁻¹; medium, incline = 9%, speed = 0.98 m·s⁻¹; high, incline = 9.0% and speed = 1.03 m·s⁻¹) chosen for each boy, which was based on his habitual activity level determined by questionnaire (Aaron et al. 1995) to encompass an appropriate range of metabolic demand for measurement of VT. The test lasted for 12 min or until a heart rate of 170 beats·min⁻¹ was achieved as previously reported (Porszasz et al. 2003).

VT was determined from ventilatory equivalent for oxygen (VE/VCO₂) and carbon dioxide (VE/VO₂) as previously

© 2008 NRC Canada
analyses, except for ICCs, which were performed using the intra-class correlation coefficient (ICC) (Streiner and Norman 1989). Adequate test–retest reliability for ICC was defined as the lower 95% confidence (ICC) (Streiner and Norman 1989). Adequate test–retest reliability for ICC was defined as the lower 95% confidence interval of repeatability (CR = 2 \times \text{standard deviation (SD)}) (Bland and Altman 1986) on the mean differences between day 1 (d1) and day 2 (d2) and the intra-class correlation coefficient (ICC) (Streiner and Norman 1989). Adequate test–retest reliability for ICC was defined as the lower 95% confidence interval of repeatability (CR = 2 \times \text{standard deviation (SD)}) (Bland and Altman 1986). Two investigators blinded to the treatment and data of the study identified the VT. The values obtained by the investigators were averaged if differences were less than 5% and re-evaluated by a third person if differences were larger than 5% (Thomas et al. 2003). Average values of the 2 determinations or the 2 closest values (in the case of 3 determinations) were used.

Estimation of fat mass

Bioelectrical impedance analysis was used to estimate body composition (RJL Systems BIA, 101Q) based on the Horlick equation (Horlick et al. 2002). The sum of skinfolds at 4 points (triceps, biceps, supra-ilial, and subscapular) was also measured by Lange skinfold caliper (Cambridge Scientific Industries, Cambridge, Md.) and recorded to the nearest 0.1 mm. The mean of 3 consecutive skinfold measurements was used for estimation of FM from an age- and sex-specific regression equation (Brook 1971). BIA was administered on 2 occasions, 1 week apart, on the same day as the FI studies. However, skinfold measurements were obtained once on the first day of the FI assessment studies and directly after the measurement of BIA.

Statistics

Reproducibility was assessed by calculating the coefficient of repeatability (CR = 2 \times \text{standard deviation (SD)}) (Bland and Altman 1986) on the mean differences between day 1 (d1) and day 2 (d2) and the intra-class correlation coefficient (ICC) (Streiner and Norman 1989). Adequate test–retest reliability for ICC was defined as the lower 95% confidence interval of repeatability (CR = 2 \times \text{standard deviation (SD)}) (Bland and Altman 1986). Two investigators blinded to the treatment and data of the study identified the VT. The values obtained by the investigators were averaged if differences were less than 5% and re-evaluated by a third person if differences were larger than 5% (Thomas et al. 2003). Average values of the 2 determinations or the 2 closest values (in the case of 3 determinations) were used.

<table>
<thead>
<tr>
<th>(d1 – d2)</th>
<th>ICC (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL (kcal)</td>
<td>–63</td>
</tr>
<tr>
<td>Water intake (g)</td>
<td>–9</td>
</tr>
<tr>
<td>Sweetness (mm)</td>
<td>3</td>
</tr>
<tr>
<td>VT (mL O2·min(^{-1}))</td>
<td>10</td>
</tr>
<tr>
<td>VT (mL O2·kg(^{-1})·min(^{-1}))</td>
<td>0.04</td>
</tr>
<tr>
<td>FM (kg)</td>
<td>–0.3</td>
</tr>
</tbody>
</table>

Table 1. Reproducibility of food and water intake, subjective sweetness, ventilation threshold, and fat mass in boys.

Note: Data are mean differences between d1 and d2 (n = 11) except for the VT (n = 8). *, p < 0.01; **, p < 0.001.

SPSS software version 15.0 (SPSS Inc.) All values are presented as means ± standard error of the mean (SEM) and p < 0.05 indicates statistical significance.

An average appetite score (measured in millimetres) was calculated at each time of measurement for each preload treatment using the following formula:

\[
\text{Appetite score} = \frac{\text{[desire to eat + hunger + (100 – fullness) + PFC]}\times 4}{100}
\]

which reflects the 4 questions on the motivation to eat VAS, as used previously (Samra et al. 2007; Bellissimo et al. 2008; Bellissimo et al. 2007a; Bellissimo et al. 2007b).

Results

Food and water intake

Lunch-time food and water intake after the glucose preload was similar when measured on 2 occasions 1 week apart (Table 1). Mean energy intake was 925 ± 139 kcal on the first day and 988 ± 147 kcal on the second day (Δ = –63 ± 39 kcal, p = 0.14). The CR, a measure of absolute agreement and the agreement relative to the variability of the measurement between subjects, showed excellent reproducibility (ICC = 0.96, p < 0.0001) (Table 1). Water intake during lunch was 229 ± 187 g on d 1 and 239 ± 185 g on d 2 (Δ = –9 ± 32 g, p = 0.77). The CR (209 g) and ICC (0.85, p < 0.01) confirm the high reproducibility of water intake at the pizza lunch (Table 1).

Subjective appetite scores and sweetness

Average appetite, desire to eat, hunger, and PFC were higher and fullness was lower on the first test visit (Fig. 2). Preload treatment (glucose d1 and glucose d2) affected average appetite (p < 0.002), desire to eat (p < 0.003), hunger (p < 0.031), fullness (p < 0.034), and PFC (p < 0.02) scores. There was also a significant main effect of time (p < 0.003) on subjective hunger, which increased to 30 min during the first test visit, but not during the second.

The CRs (range: 28–36 mm) and ICCs (range: 0.63–0.79) confirm the low reproducibility of baseline (0 min) subjec-
tive average and individual appetite scores (Table 2). Thirty minutes after the glucose preload CRs were higher and ICCs lower for desire to eat, hunger, and fullness, suggesting lower reproducibility (Table 2). However, the reproducibility was higher at 30 min for the composite average appetite score (CR = 24 mm; ICC = 0.82, p < 0.001) and PFC (CR = 13 mm; ICC = 0.94, p < 0.001).

Because absolute average and individual appetite scores differed between test days, VAS were analyzed as the change from baseline to determine whether the response to the glucose drinks was similar. The change in average appetite (main effect of preload treatment, p = 0.14), desire to eat (p = 0.34), fullness (p = 0.99), and PFC (p = 0.13) following the glucose preload did not differ between test days, but there was a significant main effect of the preload treatment (p < 0.01) for hunger ratings because the scores increased to 15 min (p < 0.004) and to 30 min (p < 0.0004) on the first test visit, but not on the second.

To determine the predictive validity of subjective appetite scores on FI, pre-lunch VAS scores for subjective appetite were correlated with the energy consumed at the pizza lunch. Thirty minutes after the preloads subjective average and individual appetite scores did not correlate with FI on either the first or second test day. Therefore, appetite scores from both days were pooled to assess the overall predictive validity of subjective appetite scores on subsequent lunchtime FI. Whole-group average appetite (r = 0.26, p = 0.24), desire to eat (r = 0.29, p = 0.19), hunger (r = 0.19, p = 0.39), fullness (r = –0.11, p = 0.61), and PFC (r = 0.35, p = 0.11) did not correlate with measured energy intake at lunch.

To assess the reproducibility of the VAS for subjective sweetness, identical glucose drinks were given to the boys on two separate occasions. Subjective sweetness of the glucose preload was not significantly different when measured on two separate occasions. Mean subjective sweetness on test d1 was 91 ± 3 mm and 88 ± 4 mm on the second test visit (Δ = 3 ± 4 mm; p = 0.38). However, the CR (23 mm) and ICC (0.36, p = 0.13) suggest there is significant variability in the ability of boys to judge sweetness (Table 1).

**Physical fitness and exercise**

The results from this study suggest that the VT is moderately reproducible in boys when measured on two separate occasions 1 week apart (Table 1). However, this conclusion is based on a sample size of only 8 boys because VT could not be determined in 3 boys on both occasions due to uninterpretable gas exchange plots. The 3 boys were excluded because of an erratic breathing pattern, which made it difficult to determine with any certainty on both occasions the break point for VT. Mean absolute VT was 990 ± 118 mL O2·min−1 on d1 and 980 ± 99 mL O2·min−1 1 week later (Δ = 10 ± 63 mL O2·min−1, p = 0.88). The CR was 359 mL·min−1 and the ICC was 0.85 (p < 0.01). VT calculated on the basis of total body weight was 22.2 ± 1.6 mL O2·kg−1·min−1 on d1 and 22.2 ± 1.3 mL O2·kg−1·min−1 1 week later (Δ = 0.04 ± 1.4 mL·kg−1·min−1, p = 0.98). A representative gas exchange plot to determine the VT from 1 subject, based on the V-slope method, is shown in Fig. 3.

Average appetite and individual appetite scores increased after a continuous, progressive, pseudo-ramp treadmill walking protocol designed to determine the VT on both occasions (Fig. 4). Paired t test comparisons showed that the increase in subjective measures of appetite after short-duration exercise was similar on both days. Subjective VAS scores for average appetite (paired t test: Δ = 0.4 ± 2.9 mm; p = 0.90), desire to eat (Δ = –6.7 ± 6.3 mm; p = 0.30), hunger (Δ = 1.1 ± 6.8 mm; p = 0.88), fullness (Δ = –2.3 ± 7.2 mm; p = 0.76), and PFC (Δ = 4.0 ± 5.3 mm; p = 0.48) did not differ between days.

Confirmation of the reproducibility of subjective appetite scores following short-duration exercise is suggested by the
Fig. 2. Appetite ratings for (a) average appetite, (b) desire to eat, (c) hunger, (d) fullness, and (e) PFC at 0, 15, and 30 min after a glucose preload on two separate days. Subjective appetite scores for average appetite \( p < 0.002 \), desire to eat \( p < 0.003 \), hunger \( p < 0.031 \), and PFC \( p < 0.015 \) were higher and fullness \( p < 0.034 \) was lower on test d2 based on PROC MIXED. *, significantly greater than 0 min by Tukey–Kramer post-hoc test \( p < 0.05 \).

Table 2. Reproducibility of subjective appetite scores before lunch in boys.

<table>
<thead>
<tr>
<th>Subjective appetite</th>
<th>(d1 – d2) Mean (mm)</th>
<th>CR (mm)</th>
<th>ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average appetite</td>
<td>–12*</td>
<td>28</td>
<td>0.73 (0.14–0.93)**</td>
</tr>
<tr>
<td>Desire to eat</td>
<td>–15*</td>
<td>36</td>
<td>0.63 (0.02–0.88)**</td>
</tr>
<tr>
<td>Hunger</td>
<td>–15*</td>
<td>34</td>
<td>0.70 (0.10–0.92)**</td>
</tr>
<tr>
<td>Fullness</td>
<td>9</td>
<td>30</td>
<td>0.79 (0.36–0.94)**</td>
</tr>
<tr>
<td>PFC</td>
<td>–9</td>
<td>33</td>
<td>0.71 (0.25–0.91)**</td>
</tr>
<tr>
<td>30 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average appetite</td>
<td>–6.2</td>
<td>24</td>
<td>0.82 (0.46–0.95)***</td>
</tr>
<tr>
<td>Desire to eat</td>
<td>–10.9</td>
<td>44</td>
<td>0.55 (0.03–0.85)*</td>
</tr>
<tr>
<td>Hunger</td>
<td>–1.6</td>
<td>41</td>
<td>0.66 (0.11–0.90)*</td>
</tr>
<tr>
<td>Fullness</td>
<td>9.9</td>
<td>52</td>
<td>0.57 (0.04–0.86)*</td>
</tr>
<tr>
<td>PFC</td>
<td>–2.4</td>
<td>13</td>
<td>0.94 (0.81–0.98)***</td>
</tr>
</tbody>
</table>

Note: Data are mean differences between d1 and d2 \( n = 11 \) for baseline (0 min) and 30 min subjective appetite scores. *, \( p < 0.05 \); **, \( p < 0.01 \); ***, \( p < 0.0001 \).
low CR for average appetite (19 mm). A Bland–Altman graph showing the mean change in average appetite ($\Delta d_1$ and $\Delta d_2$) plotted against the difference in average appetite ($d_1 - d_2$) for each subject is shown in Fig. 5. Of the individual appetite scores the CR was lowest for PFC (35 mm) and higher for desire to eat (42 mm), hunger (45 mm), and fullness (48 mm). The ICCs were low for average appetite (ICC = 0.00, $p = 0.50$), desire to eat (ICC = 0.26, $p = 0.26$), hunger (ICC = –0.15, $p = 0.66$), fullness (ICC = –0.02, $p = 0.52$), and PFC (ICC = –0.05, $p = 0.56$), because the variability between subjects was low, not because the reproducibility is low.

**Fat mass**

Fat mass estimated from BIA in boys, using the Horlick equation, is highly reproducible when measured twice, 1 week apart (Table 1). Mean FM was 8.5 ± 1.3 kg on d1 and 8.8 ± 1.2 kg on d2 ($\Delta = -0.3 \pm 0.4$ kg, $p = 0.44$). The CR was low for FM (2.7 kg) and the ICC between the two tests was high (ICC = 0.95, $p < 0.001$). However, estimates of FM from skinfolds and BIA differed in these boys. Fat mass from skinfolds (12.1 ± 1.2 kg) was significantly higher than FM estimated from BIA by 3.6 ± 0.8 kg (paired $t$ test, $p < 0.01$).

**Discussion**

The results show that lunch-time food and water intake after the glucose preload is highly reproducible and although there is significant day to day variation in subjective appetite sensations, subjective appetite scores before lunch and the increase in appetite after exercise of short duration is reproducible. The results also suggest that the VT is moderately reproducible and that the estimation of FM from BIA was highly reproducible, but underestimated FM by approximately 4 kg compared with the skinfolds measurements.

Food intake at the test lunch, 30 min after the glucose

---

**Fig. 3.** Determination of the VT by the $V_s$-slope method. A representative graph from 1 subject as obtained from an incremental load exercise test on (a) d1 and (b) d2. The VT was identified at the point where the two regression lines intersected.
preload, was reproducible when measured on 2 occasions. The CR for FI was 259 kcal and the ICC of 0.96 suggests that the absolute and relative reproducibility of FI measured after a preload is excellent in boys. The slightly high CR may have been due to the small sample size (n = 11), but it is not uncommon for studies to use small sample size to assess the reproducibility of short-term FI (Porrini et al. 1995; Raben et al. 1995). The inclusion of OW and OB boys in the analysis may have also been a factor adding variation to the CR. When the CR for FI was recalculated with only normal-weight and OW boys it was reduced to 232 kcal. In contrast, further removal of OW boys from the analysis increased the CR to 254 kcal, suggesting that OW boys, like OW hyperinsulinemic adults, tightly regulate energy intake following glucose preloads, probably owing to an adaptive response to prevent further weight gain (Samra et al. 2007).

The glucose preload did not affect the motivation to eat VAS for subjective appetite on either test visit. However, similar results for the failure of subjective appetite scores to decrease after consumption of caloric preloads have been reported in children (Anderson et al. 1989; Bellissimo et al. 2008; Bellissimo et al. 2007a; Bellissimo et al. 2007b). Despite higher subjective appetite scores on the second test day, the scores between days were similar when expressed as the change from baseline. Although the reproducibility of baseline subject appetite scores in these boys is lower than adults (Flint et al. 2000), the composite average appetite score and PFC at 30 min is highly reproducible. Furthermore, these results are consistent with other studies in adults demonstrating that mean subjective appetite scores show a similar temporal pattern of change to standardized meals on different days (Barkeling et al. 1995; Flint et al. 2000). A difference in baseline subjective appetite is expected to vary over days, but it is unclear whether it is due to true biological day to day variation of subjective appetite or methodological issues of reporting using VAS by children.

In this study, subjective appetite at 30 min did not correlate with FI, but this may have been a false negative due to the small sample size. In contrast, in all of our previous studies in boys, subjective appetite scores strongly predicted how much the boys ate at lunch irrespective of the caloric content of the preloads (Bellissimo et al. 2008; Bellissimo et al. 2007a; Bellissimo et al. 2007b). In the present study, the test meals also brought about a significant reduction in subjective appetite (data not shown), indicating that children understand the scales and can accurately express their subjective appetite using VAS. One likely explanation for the higher appetite scores on the second test visit is that the boys may have anticipated a forthcoming palatable pizza meal because of their familiarity with the study design a week earlier, thus overriding the weaker effect of the glucose preload. To test this hypothesis, future studies are re-

**Fig. 4.** Appetite ratings for average appetite, desire to eat, hunger, fullness, and PFC before and after exercise (a) d1 and (b) d2. *, p < 0.05; **, p < 0.01 by paired t test.
required where subjective appetite is monitored after larger caloric preloads in the presence and absence of a test meal.

The sweetness of caloric preloads has been proposed to contribute to the reduction of hunger and increased feelings of fullness (Lavin et al. 2002). Aspartame-sweetened drinks suppress FI in young children (Birch et al. 1989) compared with the effects of a water control. In contrast, subjective sweetness of caloric preloads was not found to be a determinant of FI in 9- to 14-year-old boys (Bellissimo et al. 2008), but this may be explained by their variability in judging sweetness. Although there was no difference in mean subjective sweetness of the glucose preloads between test days, the CR was 23 mm, meaning that a second rating by the same boy under identical conditions 1 week later may fall anywhere within a range covering 46% of the 100 mm scale.

There are no prior studies in children on the reproducibility of VT reporting the CR or ICC. In this study, absolute VT was moderately reproducible ($\Delta = 10 \text{ mL-min}^{-1}$, $p < 0.88$; $CR = 359 \text{ mL-min}^{-1}$; ICC = $0.85 p < 0.003$) when an identical fitness testing protocol was repeated on 2 separate occasions. However, the reproducibility was lower when VT was expressed on a body weight basis because the range of values for VT expressed relative to body weight was small; therefore, the ratio of within- to between-subject variation was increased. Nevertheless, other comparative studies have found that VT correlates strongly with maximum oxygen uptake in children ($r = 0.92$) (Hebestreit et al. 2000).

Why the reproducibility of the VT was only moderately reproducible is not clear, but may have been due to the small sample size ($n = 8$). Because a recent larger study in adults ($n = 40$) obtained a stronger correlation coefficient ($r = 0.94$) and a lower CR ($\sim 250 \text{ mL-min}^{-1}$), the sample size provides the likely explanation (Gaskill et al. 2001). A future study with a larger sample size (range of body weights including OW and OB children) is required to confirm the usefulness of VT as an aerobic fitness measure in children.

There are a number of practical issues that require further research before there is widespread use of VT testing in children. First, the VT is determined by visual inspection and is subject to intra- and inter-observer variability. Computerized methods to assess VT appear promising (Santos and Giannella-Neto 2004), but have not yet been tested on data from children. Second, the VT varies depending on the type of exercise or exercise protocol being performed (Davis et al. 1976). Finally, the VT cannot be determined in some subjects (approximately 20%) due to erratic breathing during the exercise protocol (Davis et al. 1976; Washington et al. 1988). However, the VT is a helpful measure of physical fitness because it does not require the child to put forth a maximal effort, it is a sensitive indicator of physical performance, and it complements VO$_2$ max in children who complain of exercise intolerance.

In addition to assessing the reproducibility of the VT, the interaction between exercise and subjective appetite in children was determined for 2 reasons. First, few studies have described the relationship between the effect of exercise and either subjective appetite or energy intake in children (Bellissimo et al. 2007b; Moore et al. 2004), although preliminary evidence from the foregoing studies suggests that age, gender, and intensity and duration of exercise are factors. Second, evidence-based physical activity recommendations for promoting healthier body weights in children are dependent on an understanding of the effect of exercise on food intake and energy balance.

Most boys experienced a net increase in subjective average appetite on both days, but a more exaggerated increase in subjective appetite was experienced by OW and OB boys.
(Fig. 5). However, the significance of the increase in subjective appetite observed in the boys after low- to moderate-intensity, short-duration exercise is uncertain because FI was not measured. The only study that has done so reported that high-intensity activity (38 min at 75% peak oxygen uptake) decreased prospective consumption immediately after in 9- to 10-year-old girls (Moore et al. 2004). These results are of interest because low-intensity, short-duration exercise programs are being promoted in many schools as a means to increase energy expenditure, but their effects on food intake have not been evaluated (Peregrin 2001).

Pediatric obesity is defined as high levels of FM that is out of proportion to other tissues (Reilly 1998). Currently there are no established guidelines for determining OW and obesity from FM estimates in children. In Canada and the US, CDC BMI charts are used for determining OW and obesity in children (Ogden et al. 2002), but the accuracy of BMI charts has been questioned (Widhalm et al. 2001). Based on the definition of obesity, simple, valid, and reproducible field methods for estimating FM are required to be used concurrently with BMI to accurately define obesity in children.

This study supports reports that BIA is a highly reproducible ($\Delta = -0.3$ kg, $CR = 2.7$ kg, ICC = 0.95) field method for estimating FM at the individual level in a pediatric population (Houtkooper et al. 1992; Houtkooper et al. 1989). Unfortunately, reproducibility of the skinfold estimates was not done because only 1 measure was made on d1 for comparison with BIA. However, skinfold measurements (mean of 3 measurements at each location used for statistical analysis) were taken by a trained observer with a technical error of $\leq 5$%; therefore, the reproducibility between days would likely have been high. There was also low agreement between BIA and skinfold estimates of FM. A recent larger study of 1254 boys and girls between 7 and 14 years of age showed similar bias such that the mean skinfold estimate of FM was 2.3% higher than BIA (Rowe et al. 2006). However, because no gold standard measure was used to validate the estimates of body composition from skinfolds and BIA, inferences about overestimation or underestimation of FM should be made with caution. Other cross-validation studies have shown that skinfold estimates of FM correlates more strongly ($r^2 = 0.62$) with a 4-compartment criterion method than BIA ($r^2 = 0.43$) (Roemmich et al. 1997). Although neither method provides an accurate estimate of FM, skinfolds seem to be better than BIA. Newer FM prediction equations are required, validated against a 4 compartment criterion method, before BIA and skinfolds can be relied upon for use at the individual level.

In summary, FI, subjective appetite after the glucose preload and short-duration exercise, and estimates of FM from BIA are highly reproducible in boys. The results also suggest that the VT is a useful and moderately reproducible measure of physical fitness in boys. Future reproducibility studies should utilize a larger sample size, including OW and OB children and girls in the analysis. In conclusion, FI, subjective appetite 30 min after the glucose preload, VT, subjective appetite after short-duration exercise, and FM from BIA were reproducible in 9- to 14-year-old boys.

Acknowledgements

Supported by an operating grant from the Canadian Institute of Health Research (grant MOP-64430) and a scholarship to Nick Bellissimo from the Danone Institute of Canada.

References


