Reproducibility and power of ad libitum energy intake assessed by repeated single meals\textsuperscript{1--3}

Nikolaj T Gregersen, Anne Flint, Christian Bitz, John E Blundell, Anne Raben, and Arne Astrup

ABSTRACT

Background: The reproducibility of the measurement of ad libitum energy intake (EI) is not well known. Furthermore, it is not known whether standardized conditions before the test day influence this measure.

Objective: The objective was to examine the reproducibility and power of the measurement of ad libitum EI with and without prior diet standardization.

Design: Fifty-five healthy, normal-weight men were tested in 2 groups, one with (D, \( n = 32 \)) and one without (ND, \( n = 23 \)) prior diet standardization, on 2 different identical occasions. They were given a standardized energy-fixed breakfast and then an ad libitum lunch 4.5 h later. Reproducibility was assessed by the coefficient of repeatability.

Results: No effect of prior diet standardization was seen on the reproducibility of ad libitum EI (\( P = 0.56 \)), but diet standardization increased ad libitum EI significantly (\( P < 0.001 \)). The correlation between ad libitum EI on the 2 test days was \( r = 0.861 (R^2 = 0.742, P < 0.0001) \) and \( r = 0.654 (R^2 = 0.428, P < 0.001) \) in the D and ND groups, respectively. The coefficient of repeatability and CV were 1478 kJ and 8.9\% compared with 1831 kJ and 14.5\% in the D and ND groups, respectively. A paired design with a study power of 0.8 requires 17 and 26 subjects, with and without prior diet standardization, respectively, to detect a difference of 500 kJ in EI.

Conclusions: The ad libitum test meal used to measure spontaneous EI is reproducible, and the reproducibility does not seem to be influenced by prior standardization. However, prior diet standardization exerts a significant effect on ad libitum EI.


INTRODUCTION

Biological markers that can predict the effect of bioactive substances on body weight and fat loss in overweight subjects are needed. A number of appetite hormones (eg, glucagon-like peptide 1) have been examined on the basis of subjective appetite recordings with the use of visual analogue scales (VAS) and ad libitum energy intake (EI) and have subsequently been shown to influence appetite and food intake (1, 2).

Novel substances and drugs with a potential influence on energy balance are frequently discovered and tested. Thus, a cost-effective and reproducible method to screen for their effects on appetite and EI is warranted. This method needs to be reproducible to be suitable for use in the current context, because this will reduce the risk of the results being unequally affected by the method, and thus make the conclusions based on this method more valid. The reproducibility, power, and validity of VAS for assessing appetite sensations in test-meal studies are well described (3–5). On the other hand, the reproducibility of EI in an ad libitum lunch has only been tested in one study, in which the authors concluded that the reproducibility of ad libitum EI was high (6). However, in this study the subjects were served a buffet-type meal for lunch, meaning that the macronutrient composition of the meals on the 2 test days varied, which could have confounded the reproducibility of the ad libitum EI because protein seems to be more satiating than carbohydrate and fat (7–10). Thus, the reproducibility of the ad libitum EI method under conditions in which the macronutrient composition is fixed has not yet been described.

In addition, it is not known whether standardized conditions before the test day have any impact on lunch intake on the subsequent day. This is relevant to clarify because, if standardization actually exerts an effect on the EI on the subsequent day, it needs to be taken into consideration when designing studies. Furthermore, it will give useful information as to whether study results can be regarded as representative of ad libitum EI under free-living conditions.

Thus, in the present article we report the reproducibility of ad libitum EI measurement in a setting in which the macronutrient composition of the meals was fixed and the power calculations needed to estimate how many subjects are needed to show a detectable difference in spontaneous EI. Conditions with or without previous diet standardization were also compared.

SUBJECTS AND METHODS

The study was conducted at the Department of Human Nutrition, Centre for Advanced Food Studies, The Royal Veterinary

\textsuperscript{1} From the Department of Human Nutrition, Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark (NTG, AF, CB, AR, and AA), and the Institute of Psychological Sciences, University of Leeds, United Kingdom (JEB).

\textsuperscript{2} Supported by The Danish Research and Development Programme for Food Technology (1995–1998) and The Danish Medical Research Council.

\textsuperscript{3} Address reprint requests and correspondence to NT Gregersen, Department of Human Nutrition, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg, Denmark. E-mail: ntg@life.ku.dk.

Received August 20, 2007.
Accepted for publication November 28, 2007.
and Agricultural University, Frederiksberg, Denmark and was approved by the Municipal Ethical Committee of Copenhagen and Frederiksberg in accordance with the Helsinki-II declaration. Results on subjective appetite sensations and palatability of the test meals were reported previously (3).

Subjects

Fifty-five healthy, normal-weight men aged 19–36 y participated in the study. None of the subjects were smokers, were elite athletes, or had a history of obesity or diabetes. The subjects were randomly assigned into 2 groups. One group (non-diet group, ND; n = 23) was instructed to abstain from physical activity and alcohol for 2 d before each test day. The other group (diet group, D; n = 32) followed the same instructions, but also consumed a standard diet for 2 d before the test days. The energy content of the standard diet was estimated according to each subject’s individual energy requirements, determined from World Health Organization tables according to age, weight, height, and sex (11). The multiplication factor used was 1.78, which corresponds to medium physical activity. The ND subjects were 26.4 (±0.5 y of age and had a body mass index (BMI) in kg/m² of 22.5 ± 0.5, whereas the D subjects were 25.0 ± 0.6 y of age and had a BMI of 22.7 ± 0.4 (NS). Their average energy requirements were estimated to be 14.0 ± 0.2 MJ/d. All subjects gave written consent after the experimental protocol had been explained to them.

Diets

Two different test meals were used: one for breakfast and one for lunch. The energy-fixed breakfast consisted of yogurt, bread, butter, cheese, jam, kiwi fruit, orange juice, and water. Total available energy content of the meal was 2 MJ for all subjects, and the energy composition was similar to that of a standard diet (50% of energy as carbohydrate, 37% of energy as fat, 13% of energy as protein, and 2.2 g/MJ dietary fiber). The lunch was a homogeneous mixed hot pot consisting of pasta, minced meat, green pepper, carrots, squash, onions, corn, and cream from which the subjects could eat ad libitum (50% of energy as carbohydrate, 37% of energy as fat, 13% of energy as protein, and 0.7 g/MJ dietary fiber).

Questionnaires

VAS, 100-mm in length with words anchored at each end, expressing the most positive and the most negative ratings were used to assess the palatability (5 questions) of the test meals. The questions were given as small booklets with one question per page. The subjects were not allowed to show or discuss their ratings with each other and were not able to refer to their previous ratings.

Statistical analyses

The mean ad libitum EIs on the 2 test days were compared by using a paired t test. A 2-sample t test with Satterthwaite adjustment of unequal variances was used to examine the effect of standardization on the reproducibility of the ad libitum EI method. Tests of reproducibility were performed according to Bland and Altman (12). Thus, the coefficient of repeatability (CR = 2 × SD) for the mean differences between meal 1 and meal 2 was calculated for the ad libitum EI, which indicates the absolute variability of the method (eg, 500 kJ), whereas the CV measures the relative variability of the method (eg, 10%). Pearson correlation analyses were used to examine correlations between ad libitum EI on the 2 test days for each group separately. Power calculations were made according to the method of Altman (13) with the use of a nomogram. For the unpaired design, the standardized difference was calculated as the difference of interest divided by the SD for the EI on test day 1 for each group separately. For the paired design, the standardized difference was calculated as twice the difference of interest divided by the SD for the within-subject difference between the EI on the 2 test days, for each group separately. For all statistical analyses, the level of significance was set to P < 0.05. The Statistical Analysis Package (version 9.1; SAS Institute, Cary, NC) was used to conduct all statistical analyses. The CV for ad libitum EI was calculated as follows:

\[ CV = \left\{ \frac{\left[ \sum(EI_1 - EI_2)^2 / (2n) \right]}{EI_{\text{mean}}} \right\}^{1/2} \]

RESULTS

No effect of prior standardization was seen on the reproducibility of the EI measurements when the subjects were pooled (P = 0.56). However, the subjects in the D group had a significantly larger ad libitum EI on both test day 1 (6010 ± 267 kJ compared with 4526 ± 221 kJ) and on test day 2 (5889 ± 221 kJ compared with 4355 ± 237 kJ) than did the subjects in the ND group (P < 0.001). Thus, it was decided to analyze the data for the 2 groups separately. No significant difference was observed in ad libitum EI for the lunch meal between the 2 test days for the D (P = 0.79) and ND (P = 0.38) groups.

The within-subject CV in the D group was 8.9%, and the correlation between EI on test days 1 and 2 was r = 0.861 (R² = 0.742, P < 0.0001; Figure 1). In the ND group the within-subject CV for EI was 14.5%, and the correlation between EI on the 2 test days was r = 0.654 (R² = 0.428, P < 0.001; Figure 2).

To illustrate the reproducibility, Bland-Altman plots for mean EI values are shown in Figure 3 (D group) and Figure 4 (ND group); the CRs are marked as horizontal lines (1478 kJ for the D group and 1831 kJ for the ND group). In Table 1 the results from the power calculations show the number of subjects needed to detect a difference in a paired or unpaired design. Calculations were made for each group separately at 2 levels of detectable differences (500 and 1000 kJ) and at 2 levels of power (0.8 and 0.9 corresponding to a 20% or 10% chance of type 2 errors, respectively). Choosing a power of 0.8 and a paired design, 17 and 26 subjects would be needed to detect a difference of 500 kJ in EI in the D and ND groups, respectively.
When using the unpaired design with 2 groups, 33 subjects per group would be needed when using prior standardization, whereas 18 subjects per group would be needed when using no prior standardization (Table 1).

The values of the different palatability scores were reported previously (3). However, in summary, the ND group rated the breakfast meal to be better tasting on the first test day (23.1 ± 2.3 mm) than on the second test day (28.2 ± 2.3 mm) \(P = 0.005\). However, apart from this, the ratings for visual appeal, smell, aftertaste, and overall palatability of the test meals were similar for the 2 different test days. Furthermore, the mean ratings for all palatability variables for both breakfast and lunch were on the positive end of the scale.

**DISCUSSION**

**Reproducibility**

We found no differences between the mean ad libitum EI on the 2 test days for either of the 2 groups. However, the CRs of 1.5 and 1.8 MJ for the D and ND groups, respectively, indicate that there was considerable individual variation in ad libitum EI between the 2 test days. Nevertheless, the test meal method seems suitable, because the variability of the ad libitum EI does not seem to depend on the size of the ad libitum EI (Figures 3 and 4). The abovementioned variation means that if a subject has an average EI of 5 MJ on a test meal, then a single test meal under identical conditions for that subject would range between 3.9 and 6.1 with prior diet standardization and between 3.7 and 6.3 MJ without prior diet standardization. This variation could be due to true biological day-to-day variation in sensations of hunger (4) and thereby ad libitum EI. However, some of the variation might...
also be caused by methodologic problems (eg, artificial conditions, slight differences in meal temperature, and room temperature—despite all attempts to make it similar), but at present it is not possible to distinguish between these 2 types of variation.

The reproducibility of the ad libitum EI method seemed higher in the D group than in the ND group ($r = 0.86$ compared with $r = 0.65$, CV $= 8.9\%$ compared with CV $= 14.5\%$, CR $= 1478$ compared with CR $= 1831$, respectively). However, reproducibility in the present study seems similar to the reproducibility in the study by Arvaniti et al (6) ($r = 0.97$, CV $= 8.2\%$), even though it might be slightly lower for the ND group in our study. Arvaniti et al did not calculate CR, but we calculated an approximate CR for that study by using the following equation:

$$\text{CR} = 2 \times \frac{(\text{CV}/100) \times \sqrt{2}}{\text{EI}_\text{mean}}$$

which becomes 1660 kJ.

$$\text{CR} = 2 \times (0.082 \times \sqrt{2}) \times \frac{(7300 + 7011)/2}{8.2}$$

The estimated CR was numerically larger than the CR for the D group but smaller than the CR for the ND group in the present study. However, reproducibility does not seem to depend on the type of meal (fixed macronutrient composition compared with buffet).

A potentially important difference between the studies was that, in our study, the macronutrient composition of the ad libitum meal was fixed, whereas in the study by Arvaniti et al (6) it varied because a buffet-type meal was used. A potential advantage of the buffet-type meal design is that the increased variety of foods (compared with the fixed meal design) might result in increased subjective ratings of meal palatability and, thereby, increased and more reliable ad libitum EIs (14). This is supported by the fact that the mean EI was $\approx 1$ and 2.5 MJ larger in the Arvaniti study than in the D and ND groups, respectively. Another potential advantage of the buffet-type meal is that it may be more representative of free-living conditions because it imposes minimal constraints on food selection and macronutrient preferences. However, an advantage of the fixed-meal design is that a potential difference in the satiating effect of macronutrients, which has been shown in some (7–10) but not all (15, 16) studies, will not affect ad libitum EI. Thus, the advantages and disadvantages of the meal types should be considered before choosing which design to use.

Another important factor that may have affected the validity of the EI measured is conversation. Although the subjects were not allowed to talk about food- and appetite-related issues, conversation on other issues may still have affected their EI, because it has been shown that distraction from a meal can affect EI, even though this effect was only seen during conversation between friends and not between strangers (17). However, if there was an effect of conversation on the EI in our study, it would presumably have been identical on both test days; thus, we do not believe that this factor significantly affected our findings.

According to the present study, diet standardization did not improve the reproducibility of the method in our subjects, but it increased the ad libitum EI on the subsequent day. The latter is a very important finding, which, to our knowledge, has not been shown before. The fact that diet standardization, even without caloric restriction, has a major impact on subsequent ad libitum EI should be considered when planning future studies. In addition, the idea of diet standardization as an improvement of the study design may need to be reconsidered, because it may elicit an eating behavior on the test day that deviates from the normal eating behavior, which may not be desirable. However, standardization of EI, alcohol consumption, and physical activity before the test day may have a positive influence on other endpoints (eg, blood glucose concentrations) by ensuring filled glycogen stores (18). Thus, further studies are needed to examine the effects of standardization on other endpoints and to examine whether our findings regarding standardization also apply to other groups of subjects.

### Power

Our results confirm that, for both groups, the number of subjects needed for a test-meal study can be reduced considerably by using a paired design. The explanation for this is that between-subject variation is eliminated when a paired design is used. However, whereas fewer subjects were needed in the D group than in the ND group when the paired design was used, the opposite was the case for the unpaired design. The reason for this was that the within-subject variation was smaller in the D group than in the ND group (SD $= 739$ kJ compared with 916 kJ), whereas the between-subject variation was larger in the D group (SD $= 1455$ kJ compared with 1061 kJ). This discrepancy in between-subject variation may have been due to the fact that the subjects were not randomly assigned to the 2 groups. However, the finding that the number of subjects needed for a test-meal study can be reduced by using a paired design is in agreement with similar analyses of VAS measurements of appetite sensations (3). According to our study, a paired design with a power of 0.8 and 0.9 will require 17 and 23 subjects, respectively, when using prior diet standardization and the outcome is EI with a detectable difference of 500 kJ (Table 1). Correspondingly, a similar setup will require 26 and 35 subjects, respectively, if no prior standardization is made (Table 1). Similarly, only 15 and 20

### TABLE 1

<table>
<thead>
<tr>
<th>Detectable difference in EI</th>
<th>Paired design</th>
<th>Unpaired design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>ND</td>
</tr>
<tr>
<td>500 kJ</td>
<td>P = 0.8</td>
<td>P = 0.9</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>1000 kJ</td>
<td>&lt;8</td>
<td>&lt;8</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

*In the unpaired design, the number of subjects refers to subjects per group.*

---

**References**

1. Arvaniti et al. (6) did not calculate CR, but we calculated an approximation of the study by Arvaniti et al (6) ($r = 0.97$, CV $= 8.2\%$), even though it might be slightly lower for the ND group in our study. Arvaniti et al did not calculate CR, but we calculated an approximate CR for that study by using the following equation:

$$\text{CR} = 2 \times \frac{(\text{CV}/100) \times \sqrt{2}}{\text{EI}_\text{mean}}$$

which becomes 1660 kJ.

$$\text{CR} = 2 \times (0.082 \times \sqrt{2}) \times \frac{(7300 + 7011)/2}{8.2}$$

The estimated CR was numerically larger than the CR for the D group but smaller than the CR for the ND group in the present study. However, reproducibility does not seem to depend on the type of meal (fixed macronutrient composition compared with buffet).

A potentially important difference between the studies was that, in our study, the macronutrient composition of the ad libitum meal was fixed, whereas in the study by Arvaniti et al (6) it varied because a buffet-type meal was used. A potential advantage of the buffet-type meal design is that the increased variety of foods (compared with the fixed meal design) might result in increased subjective ratings of meal palatability and, thereby, increased and more reliable ad libitum EIs (14). This is supported by the fact that the mean EI was $\approx 1$ and 2.5 MJ larger in the Arvaniti study than in the D and ND groups, respectively. Another potential advantage of the buffet-type meal is that it may be more representative of free-living conditions because it imposes minimal constraints on food selection and macronutrient preferences. However, an advantage of the fixed-meal design is that a potential difference in the satiating effect of macronutrients, which has been shown in some (7–10) but not all (15, 16) studies, will not affect ad libitum EI. Thus, the advantages and disadvantages of the meal types should be considered before choosing which design to use.

Another important factor that may have affected the validity of the EI measured is conversation. Although the subjects were not allowed to talk about food- and appetite-related issues, conversation on other issues may still have affected their EI, because it has been shown that distraction from a meal can affect EI, even though this effect was only seen during conversation between friends and not between strangers (17). However, if there was an effect of conversation on the EI in our study, it would presumably have been identical on both test days; thus, we do not believe that this factor significantly affected our findings.

According to the present study, diet standardization did not improve the reproducibility of the method in our subjects, but it increased the ad libitum EI on the subsequent day. The latter is a very important finding, which, to our knowledge, has not been shown before. The fact that diet standardization, even without caloric restriction, has a major impact on subsequent ad libitum EI should be considered when planning future studies. In addition, the idea of diet standardization as an improvement of the study design may need to be reconsidered, because it may elicit an eating behavior on the test day that deviates from the normal eating behavior, which may not be desirable. However, standardization of EI, alcohol consumption, and physical activity before the test day may have a positive influence on other endpoints (eg, blood glucose concentrations) by ensuring filled glycogen stores (18). Thus, further studies are needed to examine the effects of standardization on other endpoints and to examine whether our findings regarding standardization also apply to other groups of subjects.

### Power

Our results confirm that, for both groups, the number of subjects needed for a test-meal study can be reduced considerably by using a paired design. The explanation for this is that between-subject variation is eliminated when a paired design is used. However, whereas fewer subjects were needed in the D group than in the ND group when the paired design was used, the opposite was the case for the unpaired design. The reason for this was that the within-subject variation was smaller in the D group than in the ND group (SD $= 739$ kJ compared with 916 kJ), whereas the between-subject variation was larger in the D group (SD $= 1455$ kJ compared with 1061 kJ). This discrepancy in between-subject variation may have been due to the fact that the subjects were not randomly assigned to the 2 groups. However, the finding that the number of subjects needed for a test-meal study can be reduced by using a paired design is in agreement with similar analyses of VAS measurements of appetite sensations (3). According to our study, a paired design with a power of 0.8 and 0.9 will require 17 and 23 subjects, respectively, when using prior diet standardization and the outcome is EI with a detectable difference of 500 kJ (Table 1). Correspondingly, a similar setup will require 26 and 35 subjects, respectively, if no prior standardization is made (Table 1). Similarly, only 15 and 20
subjects are needed for mean appetite ratings at the same power levels (3). However, the use of a paired study design with a power of 0.8 requires 32 subjects to test both fasting and peak/nadir values for all appetite and palatability variabilities of VAS, except the desire for something sweet (n = 42), with a detectable difference of 10 mm, and to test ad libitum EI to a detectable difference of 500 kJ and a power of 0.9 (3). Thus, based on this example with detectable differences of ≈10%, it seems that the ad libitum EI method is slightly stronger than the VAS method. However, an important advantage of the VAS method is that it makes temporal tracking of appetite sensations possible, which is not possible with the ad libitum EI method. Thus, when the ad libitum EI method is used it is very important to choose the right time to test. Nevertheless, in future test-meal studies focusing on appetite and EI, it seems reasonable to use both the VAS and the ad libitum EI method, because the reproducibility of the 2 methods is moderate. Furthermore, use of both methods might provide a more detailed picture of the subject’s appetite sensations and subsequent EI, which therefore might improve the validity of the conclusions drawn. Nevertheless, more studies are needed to determine whether the ad libitum EI method is reproducible in groups that differ from the present groups in age, sex, and BMI. This is important because several factors, eg, menstrual cycle (higher EI during luteal than during follicular phase) (19), may affect the daily voluntary EI.

Taken together, despite considerable individual variations in EI for both groups, the use of an ad libitum test meal to measure spontaneous EI is reproducible. However, the validity of the conclusions based on a study using this method can be improved if VAS for appetite and palatability are also used. Furthermore, paired designs are preferred to unpaired designs when the ad libitum EI method is used. In addition, this is a cost-effective method for evaluating the appetite-regulating effect of new bioactive substances and drugs. Finally, the novel finding that diet standardization before the test day increases the ad libitum EI should be taken into consideration when planning future studies with ad libitum EI as the endpoint.

We thank Charlotte Kostecki, Karina Graff Larsen, and Berit Kristiansen for expert technical assistance in the preparation of >200 meals.

The authors’ responsibilities were as follows—AF, AR, and AA: helped write the protocol; AF: conducted the study and created the templates for the statistical analyses and the structure of the manuscript; CB, NTG, and AF: conducted the statistical analyses and helped write the manuscript; and JEB, AR and AA: helped interpret the results. All of the authors reviewed the manuscript. None of the other authors had any conflicts of interest to disclose.

REFERENCES