

ILSI Supplement

Appetite control: methodological aspects of the evaluation of foods

J. Blundell¹, C. de Graaf², T. Hulshof³, S. Jebb⁴, B. Livingstone⁵, A. Lluch⁶, D. Mela⁷, S. Salah⁶, E. Schuring⁷, H. van der Knaap⁷ and M. Westerterp⁸

¹Institute of Psychological Sciences, University of Leeds, Leeds, ²Division of Human Nutrition, Wageningen University, Wageningen, ³Kellogg's Europe, Orthen 12, Den Bosch, ⁴Nutrition and Health Research, MRC Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, ⁵School of Biomedical Sciences, University of Ulster, Coleraine, UK, ⁶Danone Research, RD 128, Palaiseau, France, ⁷Unilever R & D Vlaardingen, The Netherlands, ⁸Department of Human Biology, School of Nutrition and Metabolism, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Received 11 September 2009; revised 14 December 2009; accepted 17 December 2009

Address for correspondence: ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6, BE – 1200 Brussels, Belgium. E-mail: publications@ilsieurope.be

Summary

This report describes a set of scientific procedures used to assess the impact of foods and food ingredients on the expression of appetite (psychological and behavioural). An overarching priority has been to enable potential evaluators of health claims about foods to identify justified claims and to exclude claims that are not supported by scientific evidence for the effect cited. This priority follows precisely from the principles set down in the PASSCLAIM report. The report allows the evaluation of the strength of health claims, about the effects of foods on appetite, which can be sustained on the basis of the commonly used scientific designs and experimental procedures. The report includes different designs for assessing effects on satiation as opposed to satiety, detailed coverage of the extent to which a change in hunger can stand alone as a measure of appetite control and an extensive discussion of the statistical procedures appropriate for handling data in this field of research. Because research in this area is continually evolving, new improved methodologies may emerge over time and will need to be incorporated into the framework. One main objective of the report has been to produce guidance on good practice in carrying out appetite research, and not to set down a series of commandments that must be followed.

Keywords: Food intake hunger, satiety, measurement.

obesity reviews (2010) **11**, 251–270

Introduction

An Appetite Regulation Task Force and an expert group has been established to examine the experimental methodology on appetite control in the light of the regulatory procedures to be implemented regarding satiety and appetite claims on foods (1) and by a perceived desire to use specific foods to combat the so-called obesity epidemic. What strength of claims can be sustained on the basis of the commonly used scientific designs and experimental procedures? What constitutes a strong or weak design for assessing the effect of a food on satiety? Are different designs

required for assessing effects on satiation as opposed to satiety, and how are these related to an effect on body-weight regulation? (2,3). Does a change in hunger or satiety constitute evidence for an effect on weight management?

This report sets out the state-of-the-art in the measurement of appetite that is pertinent to the assessment of health claims in this arena. It draws together evidence to support the use of specific measures of satiation, satiety, hunger and food consumption. Research in this area is continually evolving, and new improved methodologies may emerge over time and will need to be incorporated into the framework. Additionally it is important to define those

aspects of measurement, which have clear limitations such that they are not justified or appropriate in the support of health claims. An overarching priority is to maximize the capacity to identify justified claims and to exclude claims that are not supported by scientific evidence for the effect cited. This priority follows precisely from the principles set down in the PASSCLAIM report (4). For the record, The PASSCLAIM project was an EC concerted Action on 'Process for the Assessment of Scientific Support for Claims on Foods' initiated by the EC. Through an iterative process of discussion in expert groups a workshops the project set criteria that define requirements for assessing the quality of scientific data reporting the impact of foods and food components on health and well-being. The context within which a claim can be made should be assessed by:

- considering existing legislation and dietary guidelines;
- the need for review in the light of evolving science;
- the comprehensibility of the claim to consumers.

At the outset it may be useful to have some definitions of frequently used terms to describe the operations of the appetite system. The terms may not be universally agreed, but those set out below provide a working model for discussions of methodology.

Appetite: has two definitions in circulation (i) Covers the whole field of food intake, selection, motivation and preference and (ii) Refers specifically to qualitative aspects of eating, sensory aspects or responsiveness to environmental stimulation that can be contrasted with the homeostatic view based on eating in response to physiological stimuli, energy deficit, etc.

Hunger: (i) Construct or intervening variable that connotes the drive to eat. Not directly measurable but can be

inferred from objective conditions and (ii) Conscious sensation reflecting a mental urge to eat. Can be traced to changes in physical sensations in parts of the body – stomach, limbs or head. In its strong form may include feelings of light-headedness, weakness or emptiness in stomach. Hunger will be used in this sense throughout this review.

Satiety: process that leads to the termination of eating; therefore controls meal size. Also known as intra-meal satiety.

Satiety: process that leads to inhibition of further eating, decline in hunger, increase in fullness after a meal has finished. Also known as post-ingestive satiety or inter-meal satiety.

Framework

The 'Satiety Cascade', proposed 20 years ago (5), provides a conceptual framework for examining the impact of foods on satiation (processes that bring an eating episode to an end) and satiety (processes that inhibit further eating in the postprandial period). The cascade has recently been modified by Mela (6). Macronutrient composition, energy density, physical structure and sensory qualities all contribute to the modulation of satiation and satiety. There is evidence that protein content (7) and specific viscous or gelling fibres can exert a measurable constraint over the motivation to eat; these effects depend heavily upon the amounts and types of protein and fibre used and the background food matrix, so beneficial effects of fibre and protein cannot be simply assumed (8). Of course this comment applies much more widely in appetite research, as the effect of any food or food component could be modified by the context in which it is eaten, and the effect generated

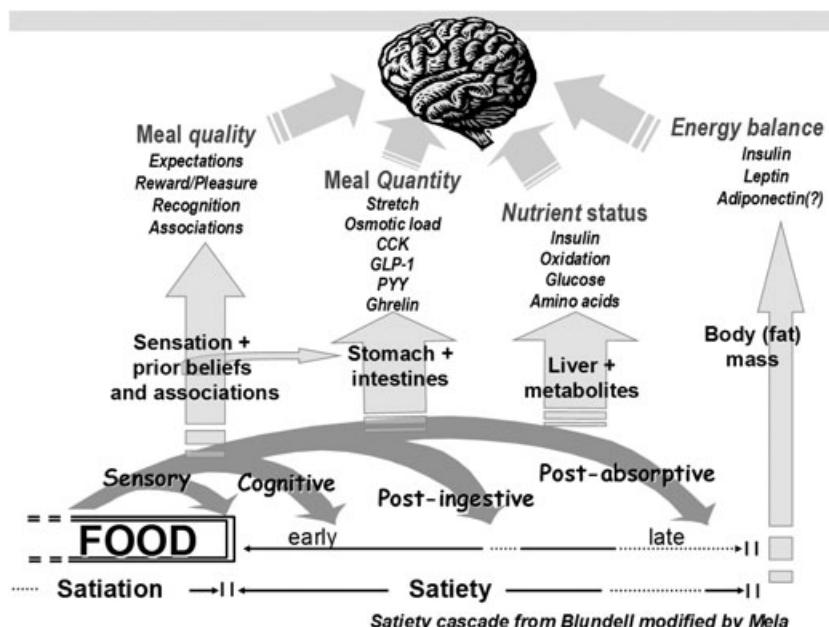


Figure 1 Example of the 'Satiety Cascade' linking the timing and sequence of eating motivations and behaviours to associated cognitive and physiological processes. CCK, cholecystokinin; GLP-1, Glucagon-like peptide; PYY, peptide YY.

by a single item tested in the laboratory may not be replicated under natural field conditions. Evidence is quite equivocal concerning the role of high and low glycaemic index foods in satiety, and the action of fat (or its subtypes) is ambiguous (9).

The biological system underlying the control of appetite is becoming better understood and involves close links between peripheral physiology and metabolism and brain processes. The system embodies mechanisms that are potential targets for foods designed to influence satiety. Of particular interest are peptides of the gastrointestinal tract, which are released following food consumption and which have physiological roles involving the management of food in the gastrointestinal tract. These include cholecystokinin, peptide YY (PYY₃₋₃₆), glucagon-like peptide (GLP-1), amylin and ghrelin.

The expert group has evaluated the strength of evidence regarding food-based mechanisms and satiety, and, where appropriate, has indicated the need for future developments. Critical processes include the balance between sensory and metabolic satiety, and between homeostatic and hedonic controls over appetite. The relationship between food palatability and satiety is critical for establishing a valid experimental methodology and for defining relevant biological targets (Fig. 1).

Strategies to measure satiation

As defined earlier, satiation is the process that brings a meal to an end. Satiation determines the size of an eating occasion. Verbal reports on the processes that bring a meal to an end indicate that 'fullness' and 'boredom with taste' are two major reasons to stop eating (10–12). These reasons may differ depending on whether we deal with the consumption of a single food or a composite meal. With a single food it is more likely that boredom of taste becomes involved, whereas with composite dishes/meals fullness may be more important in ending the meal.

Satiation is important because it determines meal size. Within the perspective of energy balance and obesity, it is instructive to note that there is no strong relationship between eating frequency and body weight (13,14). As obese people ingest more energy than non-obese people, one might infer then that meal size may be the key factor in the over consumption of energy in obese people. However, the possible contribution of meals and snacks to overeating is still a highly debated issue, which is outside the scope of this paper.

Before focusing on the methodology of measuring satiation it should be realized that in real life most meals/snacks/eating occasions are terminated through environmental factors/cues such as portion size. In most cases we finish our plate. In a field study with US soldiers on the consumption of about 5700 main meals and 8800 snacks it was

observed that soldiers ate 100% of their portion in 80–90% of the cases. Even if foods were neither liked nor disliked (a 5 on the 9-point hedonic scale), people ate on average 87% of their meal (15).

Satiation is measured through the measurement of *ad libitum* food consumption of particular experimental foods (weight in grams or energy in kcal or kJ) under standardized conditions. The *ad libitum* consumption of foods varies to a large extent. For example, a study on sensory specific satiation (16) observed that people ate on average 70–80 g from savoury cheese cookies, whereas they ate about five times as much of pears on light sugared syrup. This was not as a result of differences in liking as the pears and biscuits were about equally liked.

The proper methodology of the measurement of satiation takes into consideration those properties of food and those environmental/contextual factors that are involved in meal termination. In view of the processes of the satiety cascade (Fig. 1) it is clear that sensory factors play a major role in satiation. Many studies showed that palatability has a strong effect on *ad libitum* food intake, both from controlled experimental studies (17) as well as from more real-life studies (15). So, when studying the effect of particular food properties on satiation, it is important that the experimental foods are similarly liked.

The satiety cascade also shows that cognitive factors may play an important role in meal termination. On the basis of the consumption of many thousands of foods through our lifetime we gradually learn to estimate the satiating effects of many foods. We 'know' that we need to consume two slices of bread and cheese for breakfast in order to keep us comfortably satiated for the morning. These learning mechanisms determine our expectations about the satiating properties of foods, and probably also determine how much we put on our plate (18). This also affects how much we eat *ad libitum* from particular food in experimental situations.

A crucial role in this learned response (expectations) about satiation is the energy density of the product. In the example above where subjects consumed five times the amount in grams (weight) of pears compared with cheese biscuits, the energy density of the savoury cheese biscuits (2268 kJ per 100 g) was about eight times higher than the energy density of the pears (272 kJ per 100 g). At first sight, it seems that regular portion sizes of low-energy density foods (e.g. many liquids) are much higher than regular portion size of energy density foods (e.g. chocolate, cheese, peanuts, etc.). However, there are no systematic data in scientific papers comparing portions sizes of low-energy density foods to portion sizes of high-energy density foods. Nevertheless, this observation implies that it is crucial to match foods for energy density, when investigating the effect of food properties on satiation.

Experimental data on the effect of varying energy density on *ad libitum* food intake (with matched sensory proper-

ties) suggest that people are slow to respond to (covert) changes in energy density (19–21). This may be especially the case with beverages and foods that are eaten quickly and which do not lead to strong sensory cues (22–24). The absence of clear sensory cues may prevent people from learning to associate the oral sensory cues during eating with the post-ingestive metabolic consequences.

The texture of foods is also important in satiation. People consume more *ad libitum* from more liquid foods than they consume of more solid foods. This is related to the rate of eating that is higher in liquids than in (semi-)solids (25). The higher eating rate may be caused by the bite/swallow size that is probably higher in liquids compared with (semi-)solids. Eating 500 g of apples takes about 17 min, whereas drinking the equivalent amount of apple juice can be done in 1.5 min (26). These observations imply that controlling for texture may be an important prerequisite when investigating the effects of food properties on meal termination. However, it should be noted that slower eating does not necessarily lead to lower food intake (27).

Another cognitive factor that plays a role in meal termination is knowledge about the time until the next meal. In a study of de Graaf *et al.* (17), subjects consumed more from an *ad libitum* test meal when they knew they had no access to food the next 2 h compared with a situation where the next meal was 20 min ahead. Therefore people took into account future availability of food when deciding on their current consumption.

It is clear that the amount that people eat is also determined by the motivational state of subjects; people eat more when they hungry compared with when they are satiated. Therefore it is important that people have a similar state of satiety before presenting them with an *ad libitum* meal.

People will have certain ideas about the satiating effect of particular foods. These learned responses are mostly in concurrence with other sensory/environmental cues. When emptying a bottle of soft drinks, the weight of the bottle is a clear cue of how much we have drunk. In an elegant study, the importance of concurrent (visual) cues in meal satiation was studied by Wansink *et al.* (28). He showed that the *ad libitum* consumption of soup was strongly influenced by the visual cues concerning the emptiness of the bowl from which soup was eaten. In one condition people were (unconsciously) fooled, as the soup was refilled partially through an invisible tube. In the tube condition people ate about 73% more, while the perceived consumption was about equal. This study illustrates that people use many environmental/external cues (and not internal cues) to decide whether or not they continue eating.

In summary, the amount that we eat from a particular product is influenced by a variety of factors, related to the properties of food and the context in which the food is consumed. In general *ad libitum* consumption of foods is a

learned response based on associations between the sensory properties of foods and its metabolic consequences after ingestion. When studying the effect of properties on meal termination from a scientific point of view, it is necessary to vary one factor while holding other important factors constant. This implies that when studying satiation we need to take into consideration the palatability, energy density, and texture of foods, the motivational state (hungry vs. satiated) of subjects, potential important environmental cues (e.g. visual cues, plate size, effort to eat) and cognitive factors.

Features concerning the sensory aspects of foods on satiation and satiety

As shown in Fig. 2 food choice and intake are influenced by sensory and metabolic factors. Sensory factors are more involved in what we eat, and metabolic factors may be more involved how much we eat. Sensory signals of foods acquire their meaning/value through their association with environmental cues and through physiological, psychological and social consequences during and after eating. Within the framework of Fig. 2, there are three raw, broad concepts to consider:

- Metabolic satiation/satiety; metabolic satiety refers to all neural and hormonal signals that are transported from the gastrointestinal tract to the brain. These signals refer to stomach fullness as sensed by stretch receptors, but also to hormones involved in hunger and satiety, such as ghrelin, cholecystokinin, GLP-1 and PYY.
- Sensory specific satiation refers to the decline in reward value during consumption of a food, i.e. because of repeated exposure to a particular sensory signal. This is boredom with the taste of a particular product.
- Sensory mediated satiation/satiety signals relates to learned satiety/cephalic phase response issues; when tasting a food people know immediately something about their satiety value. This is a conditioned response based on experience with the food.

These three concepts are interrelated because of learning. Metabolic satiety concerns hunger/satiety and is responsible for energy balance, the drive to eat and to ingest enough energy. Sensory specific satiation is responsible for the drive for variety and food choice. These sensory signals from foods are linked to their metabolic consequences. That is where you get sensory mediated satiation/satiety signals. It is especially difficult to distinguish sensory specific satiation (boredom with the taste) from the learned satiation derived from repeated exposure.

When considering the effects of sensory signals on food choice and food intake, it is worthwhile to distinguish between liking and wanting. Liking is the hedonic evaluation (pleasantness, appreciation) of tasting a particular

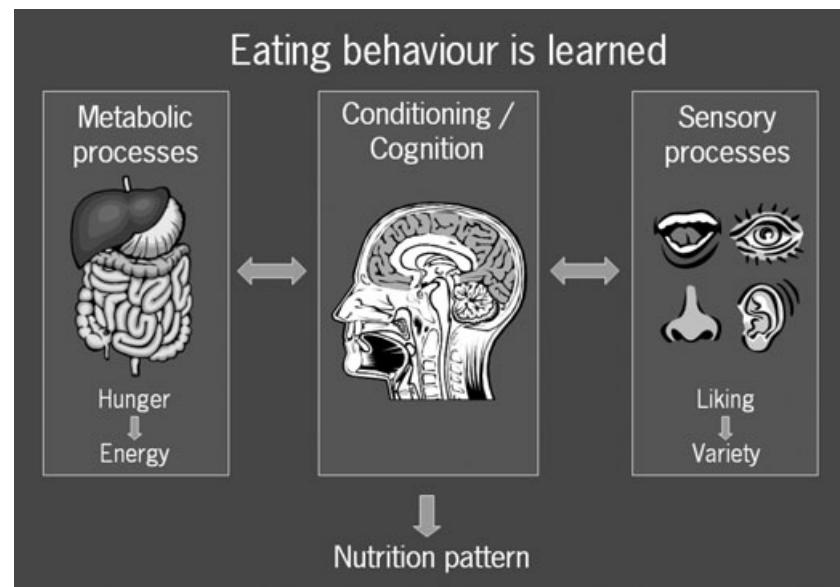


Figure 2 Model of eating behaviour. From this perspective eating behaviour is governed by three factors, metabolic factors that drive hunger and satiation, and sensory factors that drive food choice. Repeated exposure to one sensory signal will cause a drop in reward value for that particular sensory signal (see also part on sensory specific satiation within this chapter) and cause a shift to the choice for another food. In the brain the sensory signals of food are linked to the metabolic consequences, conditioning our eating/nutrition pattern. In addition, there are cognitive factors that shape our eating habits.

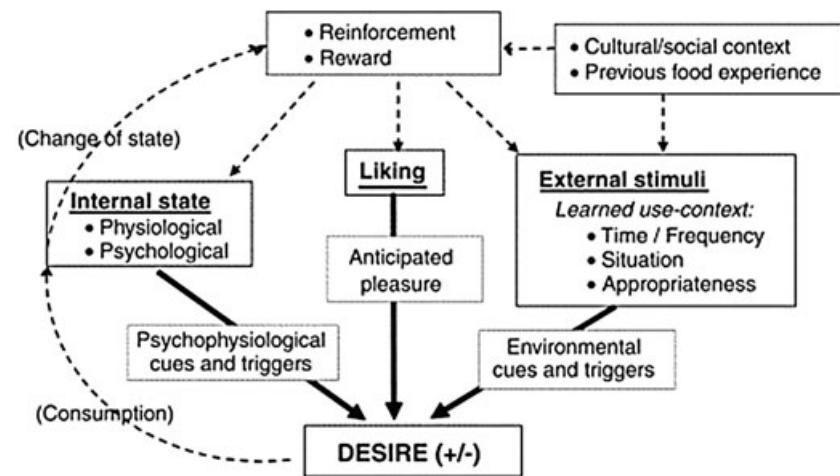


Figure 3 Scheme reflecting the effect of liking on wanting (desire) and the intermediary influences of psychophysiological state and external stimuli (learned cues). Solid lines reflect proximate drivers, dashed lines are underlying processes (Source: Mela 2006 (6)).

food, whereas wanting refers to the desire to actually ingest a particular food. The relationship between liking and wanting is well illustrated in the model developed by Mela (6). Here wanting is referred to as 'desire' to eat. This model, shown in Fig. 3, shows that liking affects wanting, but the relationship between likings and wanting is influenced by physiological/psychological state and by environmental factors. Wanting has a much more direct effect on food intake than liking. This makes the study of food wanting more relevant than the study of food liking. Recently, it has been demonstrated that liking and wanting can be experimentally dissociated. Therefore it becomes possible to examine the effects of foods, and food components, on either liking or wanting rather than on the undifferentiated hedonic response (29).

Regarding the nature of sensory signals that may have an effect on choice and intake, the sweet – savoury domain is

probably the most important dimension. This is reflected in our daily eating habits, where eating moments are often defined by the consumption of either sweet or savoury foods. The majority of energy that we ingest comes from food with either a dominant sweet taste or a dominant savoury taste. Other elements involved in sensory specific satiation are texture and flavour. After eating hard foods the desire for hard foods decline at the expense of the desire for eating softer foods (30), and a similar notion applies for different aromas or flavours.

At a more abstract level, the key sensory properties that affect sensory specific satiation (boredom) are the duration and the intensity of the sensory signal. A longer duration and a higher intensity promote satiation. Other properties that may be involved in this are the intensity and the complexity of the sensory experience. Also reinforcing properties like energy content, alcohol, caffeine may be

involved in boredom and satiation. In this respect, it must be acknowledged that there is still little hard experimental evidence on these issues. This is a major challenge for future research.

The methodology for assessing sensory specific satiation or sensory specific satiety is not well developed yet. The foods that are used to determine the effect on particular sensory properties often differ in many dimensions. This is related to the notion that it is often difficult to vary one property of a food, while holding others constant. For example, it is difficult to change the sweetness or creaminess, and not affect the energy content and other sensory properties.

Other important issues in the theories about sensory specific satiation or satiety are the bandwidth of the effects. If one gets satiated for strawberry, does one also get satiated for raspberry flavour? Does satiation or satiety generalize over different categories of foods?

One last important issue that has received very little attention until now is the quantification of the sensory experience. Although we know a lot about the relationship between sugar concentration and perceived sweetness intensity we have remarkably little insight in the strength and duration of perceived sensory intensities during eating. For example, it is clear that liquids produce shorter sensory stimulation than solids, and this may be one of the reasons for a lower satiety value sometimes reported for some simple beverages. The quantification of this observation is not available. Much more work in this field is needed to get an accurate impression of sensory influences on food intake.

Basic principle of measurement in food intake studies

Free-living vs. laboratory studies

In appetite research, the optimal experimental protocol is likely to remain elusive because of the complex and multi-faceted nature of eating behaviour. Inevitably, compromises have to be made about the requirements for internal and external validity, i.e. between precision and naturalness. In general, the internal validity of tightly controlled laboratory studies is high because they offer the highest degree of sensitivity and control over the intervention and the outcome measures. However, even when subjects are naïve to the purpose of the experiment, a combination of uncontrolled subject expectations and constraints suggests that it is highly unlikely that it is possible to fully separate the cognitive and physiological dimensions of eating behaviour under these conditions.

On the other hand, while the external validity of free-living studies is theoretically high, they are beset by a number of methodological problems that limit their inter-

nal validity. Errors in data collection are high, particularly in measurements of habitual food intake that are prone to bias, usually towards under-reporting of energy and differential mis-reporting of the macronutrients (31). Furthermore, the current difficulties in unmasking the effects of dietary components on eating behaviour under tightly controlled laboratory conditions highlight just how difficult it would be to unravel their operation in free-living circumstances. Although laboratory studies cannot replace free-living studies, they can provide crucial experimental data to complement them and it is essential that laboratory and field research in this area should advance together to help eliminate the problems inherent in both approaches and bridge the gap between them. There is clearly a lot of scope for using overlapping protocols in a variety of contexts in order that the same issues can be explored with more relevance to real-life eating behaviour and circumstances (32–35).

Preload study design

One of the most influential experimental techniques used to study the short-term regulation of food intake is the preload–test meal paradigm, carried out within part or all of a single day. Such studies are best conducted using a within subject repeated measures design and whenever possible, double-blind conditions (neither experimenter nor subject aware of the identity of the foods presented) should be observed and control conditions ensured, either by use of a non-preload or a placebo treatment. Subjects are presented with precisely prepared food(s) matched for taste, appearance and other organoleptic properties, but varying in energy and/or macronutrient composition. The hypothesis being tested will dictate if the preloads are covertly manipulated (which will assess the physiological responses to the preload) or overtly manipulated (which will test both physiological and cognitive responses) (36). After a variable time delay, the effects of the preload on spontaneous food intake are measured through accurately monitored test meal(s), or alternatively, subjects may self-report their own food intake. Subjective measures of appetite are usually taken prior to, and at predetermined time intervals after the preload and the test meal. In many of these experiments, food intake for the remainder of the day is also self-recorded by the subjects. Depending on the volume and composition of the preload and the time lapse before the test meal challenge, the respective roles of post-ingestive/pre-absorptive and post-absorptive mechanisms in the regulation of food intake can, in theory, be separated and assessed.

The apparent simplicity of, and heavy reliance on the preload experimental design has resulted in a plethora of preload studies and generated a literature that is complex, often contradictory and open to every possible interpreta-

tion. This is important when evaluating the claim for a particular food or food component. Therefore while simple in rationale, conclusions derived from preloading studies must be based on a careful scrutiny of the specific experimental conditions used. One of the major problems with these studies is that they are often deliberately designed to minimize learning about post-ingestive effects of eating, which would be expected to be highly pertinent in the real-life situation. The preload study design is also particularly prone to type 2 errors, and evidence of the sensitivity of the experimental design should always be provided (37). Factors of key importance include statistical power of the study, antecedent levels of energy deprivation and physical activity, size and composition of the preload, time lapse between the preload and test meal and test meal composition.

A number of extensions and adaptations to the preload–test meal paradigm have been made, and although these are still hostage to some of the constraints of the basic preload design, they do have enhanced external validity in so far as they simulate real-life feeding situations over longer periods of time. Some studies have manipulated the composition of both the test meals and preload challenges such that the test meal is an outcome variable in relation to the preload and the effects of both together on subsequent food intake can be evaluated (38,39).

In more medium-term studies, subjects may, or may not, reside continuously for a period of several days or a few weeks in a laboratory designed for longer-term observation of eating behaviours (14,35,40,41). They are then provided with some or all of their meals, the composition of which may be covertly manipulated, or alternatively, subjects may have relatively unrestricted access to a wide range of commercially available foods.

Preload issues

Undoubtedly, one of the main reasons why preload studies focussing on a particular issue have often generated highly variable outcomes is as a result of lack of standardization of preloads in terms of their absolute energy content, macronutrient composition, energy density, physical state (solid vs. liquid) (42–44), weight or volume (45,46) sensory (47) and cognitive (48) characteristics. Pretesting should always be done to ensure that the manipulated foods are appropriate in terms of composition, weight, volume and other sensory characteristics. Due appreciation should also be paid to the fact that eating is as much a function of the time of day at which the preload is offered and thus the ‘appropriateness’ of the food for that time of day needs to be considered.

A further key issue in this type of research is to decide whether the main dependent variable (outcome) should be the size of the next (test) meal or the time taken to the

volitional onset of the test meal. This will effectively use either size (grams of food, energy consumed) or time in minutes as the measure of the effect of the preload. Although both measures are appropriate, meal size has been the variable most frequently used in research – probably because this type of design is more convenient for the experimenter.

Timing of preload and test meal

A closely related issue is that any differential effect of the preload manipulation will diminish as the time lapse between the preload and subsequent test meal increases (49). Multiple physiological mechanisms are invoked at varying times during the post-ingestive, pre-absorptive and post-absorptive phases of satiety. However there is considerable variation between studies in this interval, ranging from <30 min (suitable for testing gastric and oro-sensory effects) to several hours (more appropriate for testing post-absorptive effects) but many researchers have failed to justify this time interval when describing their study protocols. More empirical studies that manipulate both the preload dose and the time period before the test meal are clearly required to optimize study conditions in relation to the research question being addressed. In the meantime, all study protocols should at least justify the time interval selected, otherwise decisions based largely on arbitrary criteria make it difficult to reconcile differences between study results. It is worth noting that differences in the preload–test meal interval can lead to quite different outcomes.

When the time (or latency) to next meal or inter-meal interval is used to measure satiety, it is clear that the time between the preload and the next eating episode is not fixed. In these circumstances, subjects should be time-blinded (e.g. time cues such as day-light, clocks are removed from the environment) so that meal initiation does not occur because of traditional/social time to eat. Following the preload, subjects are requested to ask for their next eating episode when feeling hungry. The time of this spontaneous meal request is recorded. Food intake at this *ad libitum* meal can also be measured (50–52).

Nature of test meal

In preload studies the most important criteria for the test meal(s) is that it/they should be sensitive to the experimental manipulations of the preload and the direction (increased or decreased intake) expected. In some studies test meals have not been offered, instead volunteers are requested to self-report their own food intakes in food diaries. However, given the dubious accuracy of self-reported intakes (31) they are no substitute for monitoring of test meal intakes under laboratory conditions. In order to ensure that voluntary food intakes are not constrained by choice or quantity, most preload studies allow subjects the opportunity to self-select from a range of normal

everyday foods, although the reasons for doing so are not always made explicit.

The composition of the test meal served after the preload can differ in the variety of foods offered. Two main options are found in the appetite research: the buffet style or the single course. In the buffet style option, a large variety of foods (including several savoury and sweet food alternatives that can differ in energy density/fat content) are offered to the subject, contrary to the single course option where the variety of foods is very limited.

The buffet style allows assessment of the potential effect of the preload manipulation on food choice (food preferences and avoidance, selection of high-energy dense foods). In principle, with this method, differences in nutrient intakes (reflecting different food choice) can be found even if energy intake remains the same between the experimental conditions.

The single course is focused on the assessment of food and energy intake rather than nutrient intakes. Therefore, the single course can be used to assess short-term energy compensation, especially when no difference in food selection is expected from the preload manipulation.

However, as noted above, despite the perceived wisdom behind it, offering a wide variety of different foods in a buffet meal scenario will not necessarily guarantee a sensitive experimental protocol as it is at variance with the usual eating pattern of the majority of people. Thus, the presence of a variety of food cues is likely to delay satiation, stimulate interest in different foods and promote increased food intake (53). Indeed, unless there is a specific hypothesis relating to food choice, then it is advisable to avoid buffet style test meals. Food selection from such meals is often prejudiced by the arbitrary nature of the foods provided and intake can be unduly influenced by items of high-energy density or salient sensory qualities. It is extremely difficult to design a selection meal in which all the items are of equal energy density, palatability and portion size.

Covert vs. overt manipulations

Manipulation of diets is a common feature of many laboratory studies, with manipulations ranging from partial alteration of the whole diet to systematic alteration of the macronutrient ratios and/or energy density of all foods in the diet. It appears that the outcomes of these studies critically depend on whether subjects only have access to covertly manipulated experimental foods or whether they have *ad libitum* access to their usual foods, outside of the experimental meals (54,55). Because experiments involving covert and highly controlled dietary manipulations are designed to minimize learning cues, the behavioural flexibility of subjects is inevitably constrained. Thus, in studies where subjects consume *ad libitum* covertly manipulated diets of constant composition, compensatory feeding

responses are likely to be blunted because subjects clearly do not have the same flexibility to alter the type, energy density and composition of foods they eat, as they would have in real life. In the absence of familiar feeding cues, the weight and volume of food is likely to attain greater significance, a phenomenon that has been observed in a number of studies that have used covertly manipulated diets over periods ranging from a few days (56–58) to a few weeks (19,59,60). These studies are characterized by a general tendency to eat a constant volume or weight of food across treatments, brought about by learned associations between the weight and volume of familiar and habitually consumed foods with the physiological consequences of ingesting those foods (61). In contrast, in the majority of studies where free choice has been accommodated alongside partial manipulation of the diet in the experimental design, more immediate and complete compensation is generally observed (32,62,63). However, an additional issue that needs to be borne in mind is that in many of these studies, variety and palatability have not been completely controlled for, although as noted above, this would be difficult to achieve in practice.

To date, most studies have favoured the covert experimental protocol but in those studies where foods have been overtly manipulated, it appears that cognitive and/or learning cues play an important role in energy compensation mechanisms (64–66). Thus in order to distinguish whether feeding responses can be attributed to the nutritional manipulation itself or to the overt or covert nature of the experimental manipulation, where possible, equal focus should be given to comparison of overt and covert manipulations using familiar foods. All of these issues clearly need to be taken on board when interpreting not only the results of such studies but also their generalizability to the free-living situation.

Duration of the study

Most experiments on the effects of foods on satiation or satiety consist of a single administration of a food on a single occasion. This is the acute food intake model. On the basis of the outcome of one single exposure of a food it is often inferred that such an effect would be observed on all future occasions if the food continued to be administered on a daily basis. However, this form of reasoning cannot be justified, and any implication that a food will continue to exert an effect with repeated administration should be accompanied by data to demonstrate this. There are a few examples in which this has been done (67).

Total energy vs. macronutrient selection – possible?

Establishing a valid food selection methodology is fraught with pitfalls. This is one reason why the approach has not

been extensively developed. The difficulty depends in large part on the degree of choice to be offered. Inviting participants to eat from a huge array of randomly assembled foods is not likely to lead to a clear understanding. A free-selection buffet is an unreliable way of measuring preferences for particular macronutrients and will reveal very little about the determinants of food selection. A strong methodology will exert control over the significant nutritional and sensory aspects of each item so that the mechanisms underlying its selection can be identified. This can be achieved most readily when choices are restricted. The option facing experimenters is whether to use a two-choice model (e.g. protein vs. carbohydrate), a three-choice model (e.g. protein, carbohydrate and fat) or a meal based on sensory–nutrient relationships (e.g. sweet or savoury taste with high or low fat).

A common tool for assessing food selection in the laboratory is the simultaneous-choice model where the independent variable (e.g. macronutrient selection) is manipulated across multiple choices within one meal (e.g. high-protein vs. high-carbohydrate foods). This model can be distinguished from the sequential-choice model where the independent variable is manipulated across two or more meals, consumed on separate days (68). The simultaneous-choice design is generally considered a reliable measure of spontaneous energy intake and food selection in the laboratory. In one within-subjects design, Arvaniti *et al.* (69) used a three-choice model repeated over two sessions to test whether participants' selections were consistent. The authors reported high correlations between amount consumed in energy ($r = 0.97$), fat ($r = 0.97$), carbohydrate ($r = 0.92$) and protein ($r = 0.82$) between the experimental sessions when separated by at least 1 week.

It follows that the food environment created in the laboratory situation will also largely determine the 'free' selection of foods. For this reason, the choices offered should be carefully controlled, while the dimensions of interest to the research objectives are systematically manipulated. Attributes that influence food selection include branding and labelling, physical properties (liquid or solid), visual appearance, micronutrient content, energy density, texture and viscosity, temperature, familiarity, palatability, sensory properties and macronutrient content. In the past, very restricted designs that permit choices between two foods were sometimes used to investigate the role of drugs on carbohydrate selection (70). However the strength of such restricted procedures depends on the protein and carbohydrate choices being matched for fat content, sweetness, palatability, etc.

An impressive, but labour intensive, method of investigating macronutrient selection in the laboratory has been developed (3). This involved a simultaneous-choice model consisting of three groups of food items rich in fat, carbohydrate or protein (~60% by energy) with the remaining energy divided between the other macronutrients to coun-

terbalance the conditions. Common, familiar, foods were included to ensure participants had some experience with the post-ingestive consequences of each food option. To control for differences in food preferences, the model involved a sizeable range of foods in each condition (10 items) such that food selection would not be compromised by avoidance of one or two foods in a particular condition.

It follows that a great deal of care and ingenuity is required to assess the selective and objective consumption of foods within test meals. Such procedures have very rarely been deployed. An alternative may be the use of computer images of foods organized according to specific categories of macronutrient and taste, for example: high fat/low fat – sweet/savoury. Using this four-compartment matrix it is possible to demonstrate clear preferences for particular categories of foods that can be modulated by prior consumption (29).

Subjective sensations as indicators of regulation

Use, reliability and validity of self-report scales in appetite research

Appetite-related self-reports include a range of measures intended to capture, over a given period, specific somatic sensations or perceived general state of hunger/repletion, motivation (desire) to eat (in general or specific food types), or prospective judgments of the quantity of food or specific foods types that could or would be eaten. The basic principles, qualities and applications of these measures in appetite studies have been concisely described by Hill *et al.* (2) and more comprehensively reviewed by Stubbs *et al.* (71) and Livingstone *et al.* (72). Thus it is only really necessary to emphasize their key points and more recent advances and evidence.

Self-report scales in common use

Specific self-report scales in common use address, for example:

- feelings of hunger, fullness, satiety;
- prospective consumption (anticipated quantity that would or could be eaten);
- desire to eat (or for a snack or meal);
- urge to eat;
- thoughts of food;
- somatic sensations (e.g. emptiness or fullness of stomach);
- desire for something sweet/savoury;
- thirst;
- nausea, gastrointestinal malaise or other side effects;
- etc.

Table 1 Recommended primary scales for self-reported appetite in healthy adults*

Scale	Question	Anchors	
		Low	High
Hunger	How hungry are you?	Not at all	Extremely As hungry as I have ever felt
Fullness	How full are you?	Not at all	Extremely As full as I have ever felt
Satiety	How satiated are you?	Not all	Extremely
Desire	How strong is your desire to eat?	Very weak Extremely low	Very strong Extremely high
Prospective consumption (quantity)	How much do you think you could (or would want to) eat right now?	Nothing at all	A very large amount

*Line scales 100–150 mm (paper) or appropriate length for electronic capture systems (see text).

In general, these scales are completed before and after consumption of the test item, and then at regular time intervals (15–30 min, up to 1 h), usually for 3–5 h, or to the start of the next meal. It is not possible to be completely prescriptive on specific scales, and translation into other dialects or languages cannot guarantee identical meaning and intensity. Furthermore, special scales would need to be developed and validated for groups such as small children, or where other perceptions may be of interest.

Nevertheless, for pragmatic reasons, and to raise the level of uniformity in the field (especially for claim support), it is desirable to recommend a standard set of basic scales for self-assessment of subjective appetite in healthy adults. The four basic scales recommended in Table 1 enjoy a history of widespread and consistent use and acceptance over several decades in many different countries and laboratories, with different test stimuli and subject groups. A good example of the use of visual analogue scales (VAS) following a standardized breakfast meal test can be found in (73). While some minor variations in terminology and anchor descriptions are possible, we recommend against significantly deviating from these unless there is strong justification to do so. Although alternatives and improvement have been proposed (74) these recommended scales are easily used and translated, and appear to be a valid, sensitive and unbiased measurement tool for regular use in standard practice. Future developments may of course lead to new alternatives.

Which scale(s) to use?

Although studies find very high, sometimes almost perfect, correlations among the different measures of general hunger/fullness, some deviations are seen in the intensity or duration, as well as sensitivity and reliability of different measures relative to each other (75,76). For this reason, it is common practice to use multiple scales. It may also be

acceptable to use a mean score comprised of several of these. However, there has been little systemic study of these points. Arguably, any of these scales could be used to support product claims, with the proviso that *wording of a claim must directly reflect the particular scale used to support it*. In other words, claims for hunger suppression should be supported by explicit measure of ‘hunger’. Furthermore, the choice of scales should be hypothesis-led: it is clearly inappropriate to use many multiple scales within a single experiment, and then selectively focus post hoc on (for example) a single scale to generate and support claims. As noted in the section on statistics, there should be a clear pre-selection of primary outcomes stated a priori in the statistical plan. Post hoc analyses of multiple scales, if this is done, require that the experiment-wide error rate be controlled, to reduce the risk of spurious ‘positive’ outcomes.

Scale types: line, category, numerical

Basic scale designs for capturing self-reports of appetite-related feelings include uni- and bipolar structured and unstructured lines, verbal categories and numerical scoring (including magnitude estimation). The most common method is the unipolar unstructured line (‘visual analogue’) scales anchored by terms such as ‘None’ or ‘Not at all’ to ‘Extreme’ or ‘As much as I have ever felt’ (Table 1).

Important scale qualities are:

- easily applied and unambiguously interpreted by investigators and subjects;
- demonstrated repeat-reliability;
- convergent validity with other, similar scales;
- known sensitivity to relevant manipulations;
- suitability for relevant mathematical and statistical handling.

Line and category scales are inherently easy to design and use, simple to explain, and require minimal data handling,

particularly if subjects can use an electronic data entry system. Although there are few direct comparisons of different scaling methods specifically in appetite research, there is a long history of research on this as applied to human perceptions and sensations in sensory psychophysics. In practice, there is probably little difference in the results from line vs. category scales, provided a sufficient number of categories are used (77). The use of electronic vs. manual scoring, and associated differences in physical line length may affect absolute scoring, but with apparently similar ability to differentiate among stimuli (71,78,79).

A theoretical problem with standard line and category scales is that scored values do not necessarily reflect perceptual distances and do not have the mathematical properties of true ratio scales. (In other words, the distance between units 1 and 2 on a line or category scale is not necessarily perceptually equivalent to the distance between units 6 and 7, nor can a score of '6' be said to be twice the perceived intensity of a score of '3'). Cardello *et al.* (74) have tested the application of a single labelled magnitude scale for satiety, which arguably overcomes these problems. Advantages are that the scale may be more sensitive and should allow for more valid quantitative transformations and comparisons. However, a major drawback is that the scale category terms must initially be selected and their scale position calibrated for each language, although it appears that if this is done with care, a single scale may be suitable to a diversity of speakers of that language (80).

Overall, the available literature suggests that different types of scales of perceived intensity largely generate quantitatively similar results for comparisons among stimuli or subject groups in various situations, although there may be advantages to a given method for particular research questions (81–85).

Reliability (repeat)

Depending upon the assessment criterion, the line scales used most commonly in appetite research (Table 1) have been found to have variable repeat-reliability with individual subjects under controlled laboratory conditions, but generally good repeat-reliability with regard to group mean data and comparisons of specific foods (71,76,78,79,86,87), even over several months or years (87). Averaged values of ratings over several hours (e.g. area under the curve [AUC]) have much better repeat-reliability than absolute scores at single time points (e.g. baseline, peak) (86,89). This also contributes to the observation that reproducibility is diminished by using change from baseline values ((87) see also statistical recommendations).

Overall, with appropriate experimental design and analysis, methods in common use have a reasonable level of reliability for comparisons of composite scores (but *not*

single time points) between foods by a given subject population. Analyses by Flint *et al.* (86) indicate that 8–35 subjects would be required to identify a 10-mm (10%) difference in 4.5-h mean appetite ratings for two foods, depending on the specific scale, level of power (0.8 or 0.9) and paired or unpaired design. This is in line with the anecdotal experience of various research groups, which suggest that under good experimental conditions, 20–25 subjects is generally sufficient to capture a 10% difference in mean or AUC appetite ratings between foods. A 10% difference is typically also seen as 'a reasonable and realistic difference' (86). Clearly more subjects are needed if there are more comparisons (to control experiment-wide error rates), or a between-subjects design. Notably, however, this robustness in response to defined manipulations probably equals or surpasses that of most associated biomarkers.

Validity (vs. other measures or constructs)

'Face validity'

There is always a fundamental difficulty in establishing the external validity of such measures (i.e. is a hunger scale really measuring 'hunger'? Is your 'hunger' the same as my 'hunger', etc.).

Rather than get tied down in this theoretical debate, a pragmatic approach can be adopted: If relevant self-report scales can effectively and reliably differentiate between stimuli in an unbiased way, in a balanced and fair test design, then they are probably suitable to substantiate the related claims.

As noted by Hill *et al.* (2), in practice these scales certainly have 'face validity': they faithfully reflect general agreement among people in certain experiences (of hunger, etc.) along a continuum, and allow for distinctions to be made within these. The scales are also generally responsive, although imperfectly, to major characteristics of meals such as volume and energy load (75,89). In addition, responses usually align in time and magnitude with corresponding physiological processes, such as stomach filling and the movement and uptake of bulk and nutrients postprandially. Lastly, under controlled conditions, there is no good evidence that variations in responses on such scales are primarily reflecting or biased towards something other than what they are purported to capture.

Self-reports as predictors of behaviour

It is important to emphasize that the validity of self-report measures is not defined by or dependent on their correlation with behavioural measures of eating (2). There are manifold reasons why self-reported appetite and actual eating may differ, and the fact that people often eat when they have low reported hunger (or vice versa), especially under free-living conditions, does not invalidate self-reports as a reflection of intensity of a specific feeling state or aspect of eating moti-

vation. With regard to product claims, demonstration of an effect on food intake is also not required. For example, where energy intake is already being effectively controlled through other means (low-energy foods, reduced-energy diet plans, meal replacements, etc.), then the benefit of enhancing satiety or reducing appetite may be simply to minimize the dysphoria of hunger feelings, independent of further changes in intake.

Nevertheless, there is a reasonable expectation that – when other factors are carefully controlled or averaged out – variation in self-reported appetite should be associated with the subsequent onset or amount of eating. Although relationships are far from perfect, this is indeed seen in both the laboratory (see reviews (2,71,72)) and real world (32,90). However, self-reported appetite is clearly not an alternative to or proxy for behavioural measures of eating. Subjective appetite and food intake measures are related, but are different consumer benefits/claim areas, which can vary independently of each other.

In short-term research studies, the relationship of perceived hunger to food intake is highly dependent upon both magnitude of differences and timing. Relative large differences in perceived hunger, especially the middle range of the scale, are most likely to relate to differences in intake. Thus, a difference in hunger ratings very late (e.g. 4–6 h) into post-meal period, when most subjects are already indicating a high level of hunger, may have much less relevance to food intake than differences earlier in the post-meal period. The latter may also be used to select the most sensitive point to test for potential differences in intakes.

Conclusions and recommendations

Although the subjective nature of self-reports may generate doubts about their use in claim substantiation, these are a sensitive, reliable and valid way to access relevant feeling states and motivations, when used within an appropriate study design and analysis.

1. Use of self-report scales (with or without related biomarkers or behavioural measures) should be strongly supported as a standard, accepted methodological approach to substantiate claims relating to effects of foods on the relevant feeling states and eating motivations.

2. Wider and more consistent use of a standard set of preferred scales in research and claim substantiation, such as those in Table 1, will improve the ability to make comparisons between research sites and studies.

3. There is a need to establish criteria for quality standards in appetite research. One approach would be a 'ring test' using an identical set of foods of fixed composition, and carefully defined subject population and methodology, to underscore the within-site reliability, sensitivity and between-site variability to detect differences between foods, using the recommended scales.

Statistical treatment of standard satiety and food intake efficacy trials

There are to date no authoritative papers or guidelines specifically addressing the appropriate statistical analysis procedures for studies of satiety and food intake. However, these studies, where satiety is measured using subjective appetite questions, may be viewed as examples of other short- or long-term clinical trials where data are collected at multiple time points. As such, they should comply with general good practice and recommended statistical standards for such trials (91–94).

Statistical analysis plan

A statistical analysis plan is part of the experiment protocol. It includes the design, the statistical analysis model, explanation of the statistical terms and the statistical results to be delivered (tables, graphs, etc.). This, along with the power determination and specific hypotheses, must be specified in detail before the execution of the trial. As a rough guide, a difference in the response magnitude relative to control in food intake or satiety ratings (AUC or overall mean) of around 8–10% is typically considered to be of practical relevance (vs. perhaps just statistical significance) (86).

The statistical plan should specify in advance which measure(s) or combination of measures (such as different rating scales) will be used as the criterion for efficacy, and powered accordingly. Conclusions based on results from one out of many different measures or rating scales requires that the experiment-wide error rate be controlled for the multiple analyses. The statistical plan should also set out any additional a priori hypotheses relating to, e.g. effects of subject characteristics (such as gender, body size, eating restraint, etc.), effects of treatments on macronutrient choice, etc. Any analyses not specified in advance, but carried out later, are by definition post hoc observations.

Design of a study

To correct for period effects and avoid introducing other artefactual effects, the experimental design should be completely balanced and randomized. To make the design balanced, a Latin square design (if possible a Williams design) should be used. For estimating treatment differences a Williams design will be sufficient to take into account a possible carry-over effect. This type of design balances the treatments over periods and subjects in such a way that each treatment is preceded by every other treatment an equal number of times. For within-subject designs, each subject will receive all treatments once, and within a period all treatments will be given to an equal number of subjects. If a carry-over effect is to be expected a strongly balanced

design is needed (for example, Luca's extra period design). These types of designs allow the estimation of both the direct effect (treatment effect) and possible extra effects like the residual effect.

Because of practical considerations (increasing duration of experiment, potential for missed test sessions, subject dropouts), it is unusual to see more than five to six treatments used in complete within-subjects designs. Where more treatments are desired within a single experiment, it is also possible to use balanced incomplete block designs, where each subject gets a subset of treatments. This will be much more efficient than a between-subjects design, while also limiting the number of treatments per subject to a feasible level.

A reduction in variance may be achieved by adjusting preload sizes or test meal intakes for individual energy requirements (based on calculated estimates or actual measures of basal metabolic rate and activity levels). This is probably preferable to post hoc analyses using measures such as body mass index or gender as indirect proxies for energy requirements. On the other hand, for testing food products sold and consumed in specific units, a fixed representative unit of that product should usually be given to all subjects.

General recommendations for main and post hoc analyses for (full-factorial) within- and also between-subject repeated measure designs

- Balanced (crossover) design (Williams or strongly balanced design).
- Because of the large between-subject variance it is usually preferred that each subject gets all treatments used in the design. Using incomplete block designs means that additional subjects are needed to maintain the same level of power.
- Experimenters should adhere to the agreed design. If a subject drops out before the study starts, replace with a new subject who gets the same treatment order.
- Number of subjects must be higher than the required (according to the power test). If there is a dropout that cannot be replaced by another subject, then there are still enough subjects to maintain the planned power.
- Each subject most preferably gets a treatment on the same day of the week, to allow 1 week of washout between two different treatments.

Analyses of self-report (e.g. satiety scores) data: single time point comparisons vs. mixed model approach (repeated measures analysis) vs. area under the curve

Although commonly done, analysis of differences in satiety or other self-report scores at multiple individual time points

(e.g. 30, 60, 90 min, etc.) is not recommended. Analysing satiety data on a single time point basis does not take into account that satiety responses are a function of multiple time points, and these individual time points are not physiologically or statistically independent. It is therefore better to use a repeated measures analysis or the AUC, which is also a continuous parameter. In addition, comparisons of many single time points leads to grossly inflated experiment-wide error rates (type 1 statistical errors), and potentially misleading interpretations and conclusions regarding efficacy.

An extra point of attention is that AUC may not be robust if the data include many missing values. In that situation it is better to use the mixed model approach with repeated measurements (RMANCOVA), which better deals with the missing data points.

Sources of variation (including their interactions) to be taken into account are listed below.

For analysis of RMANCOVA (repeated measures analysis of covariance):

- subject, the person taking the treatments;
- treatment, the compound to be tested ('wanted' effect);
- period, each treatment day;
- time, time of measure;
- time * treatment, an interaction of time and treatment;
- period * treatment, an interaction of period and treatment (carry-over and unwanted effect).

For analysis of AUC (ANCOVA):

- subject, the person taking the treatments;
- treatment, the compound to be tested ('wanted' effect);
- period, each treatment day;
- period * treatment, an interaction of period and treatment (carry-over and unwanted effect).

For higher order designs like 3 * 3, a carry-over factor (different from the period by treatment interaction) must be introduced.

Calculation of area under the curve

Areas under the curve are calculated by summarizing the mean scores of pairs of adjacent time points and then calculating a weighted mean (weighted by the time difference of two time points). For most designs, where time points are equally spaced, the weighting is not needed. Total AUC is recommended rather than the incremental or net incremental AUCs.

Calculation of duration of response

Satiety responses may differ in shape, independent of AUC, and analysis of duration of the excursion away and then

back to starting point (time to return to baseline, TTRTB) may be of interest. The difficulty that is usually encountered is that – unlike group mean responses – individual response curves rarely have a clear (inverted) *U*-shape. Thus, it is in practice difficult to identify a consistent and unique TTRTB point for individuals. Although it is also possible to plot the group mean curve and identify (by interpolation) the TTRTB, this yields only mean values, with no confidence interval. Thus, it does not allow for statistical comparisons between treatments or subject populations. Alternative approaches such as linear or non-linear regression models typically fail to adequately describe the actual shape of curves, require limiting assumptions or generate inflated variance estimates. An application of Weibull modelling to compute a mean TTRTB with confidence has recently been described (95) and offers a possibility for a consistent and practical statistical approach to this.

Change from baseline analysis

When the analysis is based on a continuous outcome there is commonly a choice of whether to use the raw outcome or changes from baseline as the primary endpoints. However, whichever of these endpoints are chosen, there is a clear consensus that the baseline value should be used as a covariate in the primary analysis in randomized studies (91–94,96,97). Analysis of covariance (ANCOVA) using the correlation between the baseline and the endpoint will increase the efficiency of the analysis vs. just analysis of change alone. Furthermore, the use of baseline subtraction ('simple change from baseline') alone, without adjusting for the baseline itself (by ANCOVA) is not generally appropriate, and can introduce artefactual effects because of random differences at baseline (whether these differences are 'statistically significant' or not).

When baseline values are included as covariates in the analyses, the estimated treatment effects are identical for analyses of both 'change from baseline' and 'raw outcome' values. Consequently if the appropriate (baseline covariate) adjustment is done, then the choice of endpoints (raw or change values) becomes solely an issue of interpretability.

Use of baseline values as covariates, rather than simple subtraction of these, is clearly also the recommended standard for evaluation of treatment effects in pharmaceutical trials (94).

Care should be taken in considering other additional covariates. In particular, in studies where food intake has already been adjusted for (estimated) individual energy requirements, this adjustment already incorporates gender, age and body weight, so those factors would not need to be entered again unless there is some additional justification. Furthermore, whereas corrections for these factors would usually be linear, relationships with energy requirements

are not, so these additional (over-)corrections may introduce errors.

Baseline comparisons

Although it is common to see, statistical testing for baseline differences has no role in a trial where the handling of the randomization has been fully satisfactory. Baseline summaries for the main covariates should be presented and discussed from a clinical point of view, irrespective of whether a statistical test indicated a 'statistically significant' difference between treatment groups or test conditions.

If the process of allocating subjects to products has, in fact, not been random then a statistically adjustment should be interpreted very cautiously. The appropriate actions (possibly excluding some subjects or centres) will follow from investigations into the cause of the imbalance.

When there is some imbalance between treatments or groups in a baseline (covariate) value that is solely as a result of chance – whether this is 'statistically significant' or not – then the appropriately adjusted analyses (ANCOVA) can account for these imbalances, whereas unadjusted analyses (including 'simple change from baseline') may not.

Quantifying satiety

Twenty-five years ago it was proposed that it would be advantageous to develop a formula for calculating the energy–satiety ratio of all common foods in order to assess their potential for causing over consumption and obesity (98). A number of procedures have been proposed to calculate the potency of foods to generate satiety; these include satiety efficiency (99), satiety index (100) and satiety quotient (101). The satiety quotient relates the suppression of hunger (or change in fullness) to the energy content of the meal consumed and has now been used in a variety of contexts that demonstrate its capacity to describe the strength of satiety generated by food consumption (102), exercise (103) or related to the biology of the individual (104,105). This formulation has some limitations mainly concerning the lack of linearity in the relationship between energy consumed and the return of hunger after eating. However, when used with fixed size meals, the formula has the capacity to provide a quantitative measure of the satiety produced by different foods.

Importance of subject variables

It is plausible to assume that not all human beings display the same uniform appetite response to a dietary manipulation or challenge. However, it may reasonably be assumed that there is an underlying common structure to the expe-

rience of appetite that includes *inter alia* eating in episodes, synchronized profiles of hunger, fullness and other sensations, and the exhibition of processes of satiation and satiety. However, the specific form of these variables and the intensity with which they are expressed are likely to be modulated by the physiological and psychological characteristics of specific groups of subjects.

Body mass index (obese vs. lean)

For a number of years, and for obvious reasons, there has been a considerable interest in describing the differences between lean and obese people in order to identify whether or not differences exist in the operation of the appetite control system and therefore in the response to particular foods or nutrients. This report does not include an extensive review of this issue. The message is that the measurement of an appetite response may vary according to the body composition or overall body mass index of an individual. It cannot be assumed that an outcome observed in a group of lean people will be replicated in a group of obese.

As an example, there is evidence that responses to pre-loads and the degree of compensation is modulated by the degree of hyperinsulinemia in male obese subjects (106). This effect has been demonstrated for passive over consumption in relation to a high-fat diet, but is likely to be true for other types of foods. This study indicates that obese subjects display blunted responses to dietary manipulations compared with lean subjects who have normal insulin levels and good insulin sensitivity. This could imply that products designed for obese (insulin resistant) people should be tested on such people. This may be important if the post-ingestive signalling system of obese subjects differs in strength from that of lean people – as indicated by the tendency to eat a constant weight/volume of foods varying in carbohydrate/fat content (107).

Although obese female subjects may show similarly shaped profiles of hunger, fullness, etc. in response to fixed size meals, when compared with lean women, the amplitude of the oscillations and the degree of satiation-induced suppression of sensations did vary (108). This indicates a difference between obese and lean women in the impact of food eaten on the subsequent feeling of hunger. It follows that claims about the effects of foods in overweight or obese people should be accompanied by evidence on these specific types of individuals.

Considering body size and body weight there is an ongoing debate about the merit of prescribing preloads or test meals as a function of body size rather than as an absolute quantity independent of size (and therefore of energy requirements). There has never been a serious examination of the effect of these two procedures, and such a study would be worth carrying out.

Psychometric traits

There is an increasing tendency to use the Three Factor Eating Questionnaire (TFEQ) or the Dutch eating Behaviour Questionnaire (DEBQ) as a screening tool in appetite studies, or as a technique to stratify subjects because of an anticipated effect on particular aspects of the appetite response. For some studies, it is now standard practice to exclude subjects with a restraint score >13 (sometimes >11). There is evidence that scores on the factors of restraint and disinhibition can influence the outcome of interventions and challenges. For example, the degree of disinhibition influences the compensatory response when the test foods vary in palatability (109). In addition, a combination of high restraint and high disinhibition can exert measurable effects (different from disinhibition alone). Subjects stratified by their restraint scores showed a different hunger and hedonic response to foods when subjected to an exercise intervention (110). There are also several examples of TFEQ factors (especially disinhibition) influencing body-weight gain (presumably by an action on feeding behaviour) (111).

Subject physiological state

Antecedent levels of energy depletion and physical activity are potentially important confounders in satiety studies. However, failure to monitor and/or standardize them is common, making it difficult to interpret differences both between and within studies. Control of antecedent diet will be particularly important in subgroups who may not be in energy balance prior to the test day (e.g. the obese and restrained eaters). If macronutrient balance is a study prerequisite, it is vital that physical activity, fasting period and alcohol intakes are standardized prior to testing in order to ensure compatibility in glycogen stores. However, not all studies involve an evaluation of energy balance. It has recently been demonstrated that the reproducibility of experimental test meals is altered very little by whether or not subjects have been standardized although the imposition of a uniform diet on the day preceding the experimental test meal (112). This is encouraging news as it suggests that the test meal scenario is fairly robust in its capacity to measure the effect of food manipulations on different occasions.

Allelic variation and individual differences

As noted above, there is now increasing recognition of the considerable heterogeneity of human populations, and consequently any nutritional or behavioural intervention can be expected to generate a wide range of responses. Under such circumstances the average response of a group of

participants may fail to disclose the diversity of responses produced. It has been noted that a 'well known phenomenon in nutritional research is the inter-individual response to any type of nutritional intervention' (113). This concept has become embodied into the field of nutritional genomics (114). The EU Framework 6 Diogenes Project has been built around the issue of characterizing the diversity in the body-weight response to different diets during a period of weight regain following a substantial enforced weight loss (115). It is inevitable that the diversity of potential responses to nutritional interventions – which have an impact on the expression of appetite – will be incorporated into the methodology of appetite research and in experimental designs for studying satiety. This implies that the possession of particular genetic polymorphisms will adjust the way in which the processes of satiation and satiety are influenced by physiology and nutrition.

Conclusions

This review of the methodology of appetite control has described the most widely used procedures together with their strengths and weaknesses. Guidance on best practice is implicit in the text. This framework provides the possibility of evaluating the strength of a claim made about the impact of a food on appetite control, hunger and satiety in the context of weight management. The impact will have been measured under a particular set of circumstances involving choices of types of test foods, amount of product administered, duration of observation period, nature of the measurement (subjective, objective) and the number of replications (if any). All of these arbitrary choices about the type of experiment carried out contribute to the strength of the claim made on behalf of a food or food component.

It is clear that, in principle, a claim can be made about the effect of a food on subjective aspects of appetite (feelings of hunger, fullness, etc.), the amount of food consumed (for example in a single test meal), the amount consumed over a whole day (24-h intake) and the effect on body weight. However, all such claims should be supported by evidence about that specific claim. If a claim is made for an effect on weight management or body weight then it is appropriate that the study has measured these specific parameters. It is inappropriate to infer that the effect of a food (or food component) on a subjective appetite feeling will be automatically translated into an effect on daily food intake, energy balance or body weight. Similar reasoning applies to a claim about weight management. It cannot be claimed that a short-term effect on hunger management will accumulate to give a long-term effect on weight management (unless body-weight data are presented). However, a legitimate and valid claim may be substantiated.

In this review it has been mentioned repeatedly that the effect of a food on satiety is determined not be some single

component within the food, but by that component interacting with other components and attributes such as energy density, active and non-active ingredients, palatability, portion size (dose), etc. Therefore the effect of a food component evaluated in isolation cannot be assumed to retain the same functional properties when administered within a food product (although this could be the case). Similarly the effect of a particular dose (quantity) of a food component may vary with the amounts of other food components with which it is eaten. These considerations simply reflect the complexity of appetite control and the difficulty of isolating the action of any single component.

It is also apparent from the above review that a specific claim on behalf of a food or food component should, in most cases, be supported by a comparison with an appropriate control food in order to judge the strength of the claim. It should be kept in mind that almost every food – including hamburgers, doughnuts and pastries, often considered responsible for over-eating – could reduce hunger (at least for a short period of time). Therefore, these foods could be claimed to be helpful – as part of a weight reducing diet. One approach to deal with this argument is to suggest that any food that carries a specific claim should have been compared with an alternative food sharing some of the same general attributes (energy value, portion size, taste, etc.) or demonstrate in other ways (for example, effects relative to its energy content) that the product and claims about it reflect a responsible positioning with regard to any implied weight management benefits. Furthermore, products carrying such claims should meet other applicable healthy eating guidelines.

Two issues have prompted this review of the methodology of appetite control: first, by the desire to use specific foods to help to combat the so-called obesity epidemic; second, by the requirement of the European Commission to legislate for the claims that are made by manufacturers about foods intended to control appetite. The widespread increase in the frequency of overweight and obesity is evidence of the ease with which food consumption (relative to a low-energy expenditure) can lead to a positive energy balance and weight gain. The capacity of human beings to gain weight incrementally over time is very difficult to halt and even more difficult to reverse. Sometimes this can be achieved by control of the total diet or by a major permanent shift in eating pattern. Therefore to produce and identify single foods or food components that can impact on daily energy intake, energy balance or body weight constitutes a major challenge. Some functional foods for appetite control will have the capacity to achieve positive effects over a particular time period. The identification of these foods with certainty for the benefit of the consumer requires the application of a strong scientific methodology as envisaged in the PASSCLAIM principles. The essential features of this methodology have been outlined in this report.

In conclusion, it is worth setting out a few principles that can be kept in mind when designing, or evaluating, studies in the field of appetite control:

1. PASSCLAIM principles apply, with additional considerations for this specific claim area;
2. No single test of appetite that applies to all claims;
3. Clarity of claim and data required to support the specific claim;
4. Claims regarding appetite/satiety are not claims about weight control *per se* and are not substantiation for weight control claims;
5. Validated and appropriate analysis (as described in this report and checklist);
6. Biomarker and mechanistic data are not adequate substantiation for a claim in the absence of primary substantiation by behavioural measures.

The following checklist can also be used

1. Refer to PASSCLAIM to identify other generic issues;
2. Statistical plan and analyses clear;
 - Primary hypotheses (including subgroup analyses) pre-specified;
 - Adequately powered (within-subject preferred);
 - Balanced treatment order, random allocation;
 - Energy matched treatments;
 - Treatment effect statistically compared with control effect;
 - Appropriate handling of baseline values (ANCOVA recommended);
 - Within-experiment error rate controlled for multiple testing.
3. Population reflects intended target (product user) group;
4. Appropriately matched control products (sensory, nutritional, etc.) – ‘1 variable principle’;
5. Food intake claims supported by objective measures of intake;
6. Test product or ingredient representative of product/ingredient as eaten and carrying claim;
 - Dose level, matrix;
 - Material (ingredient) specification and handling, processing and storage.
7. Meaningful effect size (clinical vs. statistical significance);
8. *In vitro* data, biomarkers, potential mechanisms supported by direct (behavioural) evidence.

Conflict of Interest Statement

The authors state that there are no conflicts of interest.

Acknowledgements

This work was commissioned by the Appetite Regulation Task Force of the European branch of the International Life Sciences Institute (ILSI Europe). Industry members of this task force are: Ajinomoto Europe, Cadbury, Cargill, Coca-Cola Europe, Colloïdes Naturels International, Cosucra, Danisco, Groupe Danone, Kellogg Europe, Kraft Foods, Lipid Nutrition, National Starch Food Innovation, Nestlé, PepsiCo International, Sensus, Südzucker/BENE Group, Tate & Lyle and Unilever. For further information about ILSI Europe, please email info@ilsieurope.be or call +32 2771 00 14. The opinions expressed herein are those of the authors and do not necessarily represent the views of ILSI Europe. The work of the expert group and its publication was coordinated by Fiona Samuels, Scientific Project Manager at ILSI Europe.

References

1. Riccardi G, Aggett P, Brighenti F, Delzenne N, Frayn K, Nieuwenhuizen A, Pannemans D, Theis S, Tuijtelars S. PASSCLAIM-body weight regulation, insulin sensitivity and diabetes risk. *Eur J Nutr* 2004; 43(Suppl. 2): II7-II46.
2. Hill AJ, Rogers PJ, Blundell JE. Techniques for the measurement of human eating behaviour and food intake: a practical guide. *Int J Obesity* 1995; 19: 361-375.
3. Stubbs RJ, O'Reilly LM, Johnstone AM, Harrison CLS, Clark H, Franklin MF, Reid CA. Description and evaluation of an experimental model to examine changes in selection between high-protein, high-carbohydrate and high-fat foods in humans. *Eur J Clin Nutr* 1999; 53: 13-21.
4. Aggett PJ, Antoine JM, Asp NG, Bellisle F, Contor L, Cummings JH, Howlett J, Muller DJ, Persin C, Pijls LT, Reckemmer G, Tuijtelars S, Verhagen H. PASSCLAIM: consensus on criteria. *Eur J Nutr* 2005; 44(Suppl. 1): i5-i30.
5. Blundell JE, Rogers PJ, Hill AJ. Evaluating the satiating power of foods: implications for acceptance and consumption. In: Solms J, Booth DA, Pangbourne RM, Raunhardt O (eds). *Food Acceptance and Nutrition*. Academic Press: London, 1987, pp. 205-219.
6. Mela DJ. Eating for pleasure or just wanting to eat? Reconsidering sensory hedonic responses as a driver of obesity. *Appetite* 2006; 47: 10-17.
7. Lejeune MPM, Westerterp KR, Adam TCM, Luscombe-March ND, Westerterp-Platenga MS. Ghrelin and glucagon-like peptide 1 concentrations, 24-h satiety, and energy and substrate metabolism during a high protein diet and measured in a respiration chamber. *Am J Clin Nutr* 2006; 83: 89-94.
8. Slavin JL. Dietary fibre and body weight. *Nutrition* 2005; 21: 411-418.
9. Lawton CL, Delargy HJ, Brockman J, Smith FC, Blundell JE. The degree of saturation of fatty acids influences post-ingestive satiety. *Br J Nutr* 2000; 83: 473-482.
10. Hetherington MM, Macdiarmid JI. Pleasure and excess: liking for and overconsumption of chocolate. *Physiol Behav* 1995; 57: 27-35.
11. Mook DG, Votaw MC. How important is hedonism? Reasons given by college students for ending a meal. *Appetite* 1992; 18: 69-75.
12. Tuomisto T, Tuomisto MT, Hetherington M, Lappalainen R. Reasons for initiation and cessation of eating in obese men and

women and the affective consequences of eating in everyday situations. *Appetite* 1998; 30: 211–222.

13. Bellisle F, McDevitt R, Prentice AM. Meal frequency and energy balance. *Br J Nutr* 1997; 77: S57–S70.
14. Johnstone AM, Shannon E, Whybrow S, Reid CA, Stubbs RJ. Altering the temporal distribution of energy intake with isoenergetically dense foods given as snacks does not affect total daily energy intake in normal-weight men. *Br J Nutr* 2000; 83: 7–14.
15. de Graaf C, Cardello AV, Kramer M, Lesher LL, Meiselman HL, Schutz HG. A comparison between liking ratings obtained under laboratory and field conditions: the role of choice. *Appetite* 2005; 44: 15–22.
16. Weenen H, Stafleu A, de Graaf C. Dynamic aspects of liking: post-prandial persistence of sensory specific satiety. *Food Qual Pref* 2005; 16: 528–535.
17. de Graaf C, de Jong LS, Lambers AC. Palatability affects satiation but not satiety. *Physiol Behav* 1999; 66: 681–688.
18. Brunstrom J. Associative learning and the control of human dietary behaviour. *Appetite* 2007; 49: 268–271.
19. Lissner L, Levitsky DA, Strupp BJ, Kalkwarf HJ, Roe DA. Dietary fat and the regulation of energy intake in human subjects. *Am J Clin Nutr* 1987; 46: 886–892.
20. Kendall A, Levitsky DA, Strupp BJ, Lissner L. Weight loss on a low-fat diet: consequence of the imprecision of the control of food intake in humans. *Am J Clin Nutr* 1991; 53: 1124–1129.
21. Blundell JE. Control of human appetite: implications for the intake of dietary fat. *Annu Rev Nutr* 1996; 16: 285.
22. Raben A. Should obese patients be counselled to follow a low-glycaemic index diet? No. *Obes Rev* 2002; 3: 245.
23. DiMeglio DP, Mattes RD. Lipid versus solid carbohydrate: effects on food intake and body weight. *Int J Obes Relat Metab Disord* 2000; 24: 794–800.
24. Ebbeling CB, Sinclair KB, Pereira MA, Garcia-Lago E, Feldman HA, Ludiwg DS. Compensation for energy intake from fast food among overweight and lean adolescents. *JAMA* 2004; 291: 2828–2833.
25. Zijlstra N, Mars M, Stafleu A, de Wijk RA, Prinz JF, Hück NL, de Graaf C. Effect of bite size and oral processing time of food on satiation. *Appetite* 2008; 51: 753–753.
26. Haber GB, Heaton KW, Murphy D, Burroughs LF. Depletion and disruption of dietary fibre: effects on satiety, plasma-glucose, and serum-insulin. *Lancet* 1977; 310: 679–682.
27. Martin CK, Anton SD, Walden H, Arnett C, Greenway FL, Williamson DA. Slower eating rate reduced the food intake of men, but not women: implications for behavioural weight control. *Behav Res Ther* 2007; 45: 2349–2359.
28. Wansink B, Painter J, North J. Bottomless bowls: why visual cues of portion size may influence intake. *Obesity Res* 2005; 13: 93–100.
29. Finlayson GS, King NA, Blundell JE. Liking vs. wanting food: importance for human appetite control and weight regulation. *Neurosci Biobehav Rev* 2007; 31: 987–1002.
30. Guinard J-X, Brun P. Sensory specific satiety: comparison of taste and texture effects. *Appetite* 1998; 31: 141–157.
31. Livingstone MBE, Black AE. Markers of the validity of reported energy intake. *J Nutr* 2003; 133: 895S–920S.
32. Foltin RW, Fischman MW, Moran TH, Rolls BJ, Kelly TH. Caloric compensation for lunches varying in fat and carbohydrate content by humans in a residential laboratory. *Am J Clin Nutr* 1990; 52: 969–980.
33. de Castro JM. Varying levels of food energy self-reporting are associated with between-group, but not within-subject, differences in food intake. *J Nutr* 2006; 136: 1382–1388.
34. de Castro JM, Bellisle F, Dalix A-M, Pearcey SM. Palatability and intake relationships in free-living human: characterization and independence of influence in North Americans. *Physiol Behav* 2000; 70: 343–350.
35. Rumpler WV, Kramer M, Rhodes DG, Paul DR. The impact of the covert manipulation of macronutrient intake on energy intake and variability in daily food intake in nonobese men. *Int J Obes* 2006; 30: 774–781.
36. Rolls BJ, Hammer VA. Fat, carbohydrate, and the regulation of energy intake. *Am J Clin Nutr* 1995; 62(Suppl.): 1086S–1095S.
37. Blundell JE, Stubbs RJ. Diet composition and the control of food intake in humans. In: Bray GA, Bouchard C, James WPT (eds). *International Handbook of Obesity*. Dekker Inc.: New York, 1998, pp. 243–272.
38. Lawton CL, Burley VJ, Wales JK, Blundell JE. Dietary fat and appetite control in obese subjects: weak effects on satiation and satiety. *Int J Obes Relat Metab Disord* 1993; 17: 409–416.
39. Green SM, Wales JK, Lawton CL, Blundell JE. Comparison of high-fat and high-carbohydrate foods in a meal or snack on short-term fat and energy intakes in obese women. *Br J Nutr* 2000; 84: 521–530.
40. Foltin R, Rolls B, Moran T, Kelly T, McNeilis A, Fischman M. Caloric, but not macronutrient, compensation by humans for required-eating occasions with meals and snacks varying in fat and carbohydrate. *Am J Clin Nutr* 1992; 55: 331–342.
41. Alfenas RCG, Mattes RD. Influence of glycemic index/load on glycemic response, appetite, and food intake in healthy humans. *Diabetes Care* 2005; 28: 2123–2129.
42. Almiron-Roig E, Chen Y, Drewnowski A. Liquid calories and the failure of satiety: how good is the evidence? *Obes Rev* 2003; 4: 201–212.
43. Mattes R. Soup and satiety. *Physiol Behav* 2005; 83: 739–747.
44. Mourao DM, Bressan J, Campbell WW, Mattes RD. Effects of food form on appetite and energy intake in lean and obese adults. *Int J Obes Relat Metab Disord* 2007; 31: 1688–1695.
45. de Graaf C, Hulshof T. Effects of weight and energy content of preloads on subsequent appetite and food intake. *Appetite* 1996; 26: 139–151.
46. Rolls BJ, Castellanos VH, Halford JC, Kilara A, Panyam D, Pelkman CL, Smith GP, Thorwart ML. Volume of food consumed affects satiety in men. *Am J Clin Nutr* 1998; 67: 1170–1177.
47. Sorensen LB, Moller P, Flint A, Martens M, Raben A. Effect of sensory perception of foods on appetite and food intake: a review of studies on humans. *Int J Obes Relat Metab Disord* 2003; 27: 1152–1166.
48. Pliner P, Zec D. Meal schemas during a preload decrease subsequent eating. *Appetite* 2007; 48: 278–288.
49. Rolls BJ, Kim S, McNeilis AL, Fischman MW, Foltin RW, Moran TH. Time course of effects of preloads high in fat or carbohydrate on food intake and hunger rating in humans. *Am J Physiol* 1991; 260: R756–R763.
50. Marmonier C, Chapelot D, Louis-Sylvestre J. Effects of macronutrient content and energy density of snacks consumed in a satiety state on the onset of the next meal. *Appetite* 2000; 34, 161–168.
51. Melanson KJ, Westerterp-Plantenga MS, Saris WH, Smith FJ, Campfield LA. Blood glucose patterns and appetite in time-blinded humans: carbohydrate versus fat. *Am J Physiol* 1999; 277(2 Pt 2): R337–R345.
52. Himaya A, Fantino M, Antoine JM, Brondel L, Louis-Sylvestre J. Satiety power of dietary fat: a new appraisal. *Am J Clin Nutr* 1997; 65: 1410.

53. Hetherington MM, Foster R, Newman T, Anderson AS, Norton G. Understanding variety: tasting different foods delays satiation. *Physiol Behav* 2006; **87**: 263–271.

54. Stubbs RJ, Mullen S, Johnstone AM, Rist M, Kracht A, Reid C. How covert are covertly manipulated diets? *Int J Obes Relat Metab Disord* 2001; **25**: 567–573.

55. Poppitt SD, Prentice AM. Energy density and its role in the control of food intake: evidence from metabolic and community studies. *Appetite* 1996; **26**: 153–174.

56. Rolls BJ, Bell EA, Thorwart ML. Water incorporated into a food but not served with a food decreases energy intake in lean women. *Am J Clin Nutr* 1999; **70**: 448–455.

57. Araya H, Vera G, Alvena M. Effect of the energy density and volume of high carbohydrate meals on short term satiety in pre-school children. *Eur J Clin Nutr* 1999; **53**: 273–276.

58. Bell EA, Rolls BJ. Energy density of foods affects energy intake across multiple levels of fat content in lean and obese women. *Am J Clin Nutr* 2001; **73**: 1010–1018.

59. Thomas CD, Peters JC, Reed GW, Abumrad NN, Ming S, Hill JO. Nutrient balance and energy expenditure during ad libitum feeding of high-fat and high-carbohydrate diets in humans. *Am J Clin Nutr* 1992; **55**: 934–942.

60. Stubbs RJ, Harbord CG, Murgatroyd PR, Prentice AM. Convert manipulation of dietary fat and energy density: effect on subsequent flux and food intake in men feeding ad libitum. *Am J Clin Nutr* 1995; **62**: 316–330.

61. Blundell JE. The psychobiological approach to appetite and weight control. In: Brownell KD, Fairburn CG (eds). *Eating Disorders and Obesity: A Comprehensive Handbook*. The Guildford Press: Guildford, 1995, pp. 13–20.

62. Mattes RD, Pierce CB, Friedman MI. Daily caloric intake of normal weight adults: response to changes in dietary energy density of a luncheon meal. *Am J Clin Nutr* 1988; **48**: 214–219.

63. Foltin RW, Fischman MW, Emurian CS, Rachlinski JJ. Compensation for caloric dilution in humans given unrestricted access to food in a residential laboratory. *Appetite* 1988; **10**: 13–24.

64. Caputo FA, Mattes RD. Human dietary responses to perceived manipulation of fat content in a midday meal. *Int J Obes Relat Metab Disord* 1993; **17**: 237–240.

65. Aaron JI, Mela DJ, Evans RE. The influence of attitudes, beliefs and label information on perceptions of reduced-fat spreads. *Appetite* 1994; **22**: 25–37.

66. Shide DJ, Rolls B. Information about the fat content of pre-loads influences energy intake in healthy women. *J Am Diet Assoc* 1995; **95**: 993–998.

67. King SC, Weber AJ, Meiselman Herbert L, Nan LV. The effect of meal situation, social interaction, physical environment and choice of food acceptability. *Food Qual Pref* 2004; **15**: 645–653.

68. Blundell JE, Stubbs RJ, Golding C, Croden F, Alam R, Whybrow S, LeNoury J, Lawton CL. Resistance and susceptibility to weight gain: individual variability in response to a high fat diet. *Physiol Behav* 2005; **86**: 614–622.

69. Arvanitaki K, Richard D, Tremblay A. Reproducibility of energy and macronutrient intake and related substrate oxidation rates in a buffet-style meal. *Br J Nutr* 2000; **83**: 489–495.

70. Wurtman J, Wurtman R, Reynolds S, Tsay R, Chew B. Fenfluramine suppresses snack intake among carbohydrate cravers but not among non-carbohydrate cravers. *Int J Eat Disord* 1987; **6**: 687–699.

71. Stubbs RJ, Hughes DA, Johnstone AM, Rowley E, Reid C, Elia M, Stratton R, Delargy H, King N, Blundell JE. The use of visual analogue scales to assess motivation to eat in human subjects: a review of their reliability and validity with an evaluation of new hand-held computerized systems for temporal tracking of appetite ratings. *Br J Nutr* 2000; **84**: 405–415.

72. Livingstone MBE, Robson PF, Welch RW, Burns AA, Burrows MS, McCormack C. Methodological issues in the assessment of satiety. *Scand J Nutr/Näringsforskning* 2000; **44**: 98–103.

73. Raben A, Christensen NJ, Madsen J, Holst JJ, Astrup A. Decreased post prandial thermogenesis and fat oxidation but increased fullness after a high fibre meal compared with a low fibre meal. *Am J Clin Nutr* 1994; **59**: 1386–1394.

74. Cardello AV, Schutz HG, Lesher LL, Merrill E. Development and testing of a labeled magnitude scale of perceived satiety. *Appetite* 2005; **44**: 1–13.

75. Porrini M, Crovetto R, Testolin G, Silva S. Evaluation of satiety sensations and food intake after different preloads. *Appetite* 1995; **25**: 7–30.

76. Merrill EP, Kramer FM, Cardello A, Schutz H. A comparison of satiety measures. *Appetite* 2002; **39**: 181–183.

77. Jeon S-Y, O'Mahony M, Kim K-O. A comparison of category and line scales under various experimental protocols. *J Sens Stud* 2004; **19**: 49–66.

78. Stubbs RJ, Hughes DA, Johnstone AM, Rowley E, Ferris S, Elia M, Stratton R, King N, Blundell JE. Description and evaluation of a Newton-based electronic appetite rating system for temporal tracking of appetite in human subjects. *Physiol Behav* 2001; **72**: 615–619.

79. Stratton RJ, Stubbs RJ, Hughes D, King N, Blundell JE, Elia M. Comparison of the traditional paper visual analogue scale questionnaire with an Apple Newton electronic rating system (EARS) in free living subjects feeding ad libitum. *Eur J Clin Nutr* 1998; **52**: 737–741.

80. Zalilah MK, Greenway DR, Caffin NA, D'Arcy BR, Gidley MJ. Application of labelled magnitude satiety scale in a linguistically-diverse population. *Food Qual Pref* 2008; **19**: 574–578.

81. Giovanni MA, Pangborn RM. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. *J Food Sci* 1983; **48**: 1175–1182.

82. Lawless HT. Logarithmic transformation of magnitude estimation data and comparison of scaling methods. *J Sens Stud* 1989; **4**: 75–86.

83. Lawless HT, Horne J, Spiers W. Contrast and range effects for category, magnitude and labeled magnitude scales in judgements of sweetness intensity. *Chem Senses* 2000; **25**: 85–92.

84. Lawless HT, Malone GJ. The discriminative efficiency of common rating scaling. *J Sens Stud* 1986; **1**: 85–98.

85. Lawless HT, Malone GJ. A comparison of rating scales: sensitivity, replicates and relative measurement. *J Sens Stud* 1986; **1**: 155–174.

86. Flint A, Raben A, Blundell JE, Astrup A. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int J Obes* 2000; **24**: 38–48.

87. Raben A, Tagliabue A, Astrup A. The reproducibility of subjective appetite scores. *Br J Nutr* 1995; **73**: 517–530.

88. Kruse H-P. Can satiety be measured? *Nahrung/Food* 2001; **45**: 298–301.

89. Kirkmeyer SV, Mattes RD. Effects of food attributes on hunger and food intake. *Int J Obes* 2000; **24**: 1167–1175.

90. Mattes R. Hunger ratings are not a valid proxy measure of reported food intake in humans. *Appetite* 1990; **15**: 103–113.

91. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; **21**: 2917–2930.

92. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355**: 1064–1069.

93. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med* 2006; **25**: 4334–4344.

94. The Medicines Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products. Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99, 2003. [WWW document]. URL <http://www.ema.europa.eu/pdfs/human/referral/naftidofuryl/05789Sen.pdf> (accessed 14 January 2010).

95. Kovacs EMR, Schuring E, Quadt F, Wiseman S, Mela DJ. Measurement and comparison of duration of satiety responses: a new quantitative procedure. *Obesity* 2007; **15**(Suppl.): A155.

96. Overall JE, Ashby B. Baseline corrections in experimental and quasi-experimental clinical trials. *Neuropsychopharmacology* 1991; **4**: 273–281.

97. van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies. *J Clin Epidemiol* 2006; **59**: 920–925 (Erratum: *J Clin Epidemiol* 2006; **59** (12): 1334).

98. Heaton KW. Dietary fibre and energy intake. In: Berchtold P, Cairella A, Jacobelli A, Silano V (eds). *Regulators of Intestinal Absorption in Obesity, Diabetes and Nutrition*. Societa Editrice Universo: Rome, 1981, pp. 283–294.

99. Kissileff HR. Satiating efficiency of a strategy for conducting food loading experiments. *Neurosci Biobehav Rev* 1984; **8**: 129–135.

100. Holt SH, Miller JC, Petocz P, Farmakalidis E. A Satiety Index of common foods. *Eur J Clin Nutr* 1995; **49**: 675–690.

101. Green SM, Delargy HJ, Joanes D, Blundell JE. A satiety quotient: a formulation to assess the satiating effect of food. *Appetite* 1997; **29**: 291–304.

102. Parker BA, Sturm K, MacIntosh CG, Feinle C, Horowitz M, Chapman IM. Relation between food intake and visual analogue scale ratings of appetite and other sensations in healthy older and young subjects. *Eur J Clin Nutr* 2004; **58**: 212–218.

103. King NA, Hopkins M, Caudwell P, Stubbs RJ, Blundell JE. Individual variability following 12 weeks of supervised exercise: identification and characterisation of compensation for exercise. *Int J Obes* 2008; **32**: 177–184.

104. Drapeau V, Blundell JE, Therrien F, Lawton CL, Richard D, Tremblay A. Appetite sensations as markers of overall intake. *Br J Nutr* 2005; **93**: 273–280.

105. Drapeau V, King N, Hetherington M, Doucet E, Blundell JE, Tremblay A. Appetite sensations and the satiety quotient: predictors of energy intake and weight loss. *Appetite* 2007; **48**: 159–166.

106. Speechley DP, Buffenstein R. Appetite dysfunction in obese males: evidence for role of hyperinsulinaemia in passive overconsumption with a high fat diet. *Eur J Clin Nutr* 2000; **54**: 225–233.

107. Blundell JE, Cotton JR, Delargy HJ, Green SM, Greenough A, King N, Lawton CL. The fat paradox: fat-induced satiety signals versus high fat over consumption. *Int J Obes* 1995; **19**: 832–835.

108. Barkeling B, King N, Naslund E, Blundell JE. Characterisation of individuals who claim to detect no relationship between their eating pattern and sensations of hunger and fullness. *Int J Obes* 1998; **31**: 435–439.

109. Yeomans MR, Tovey HM, Tinley EM, Haynes CJ. Effects of manipulated palatability on appetite depend on restraint and disinhibition scores from the Three-Factor Eating Questionnaire. *Int J Obes Relat Metab Disord* 2004; **28**: 144–151.

110. Lluch A, King N, Blundell JE. Exercise in dietary restrained women: no effect on energy intake but change in hedonic ratings. *Eur J Clin Nutr* 1998; **52**: 300–307.

111. Bryant EJ, King NA, Blundell JE. Disinhibition: its effects on appetite and weight regulation. *Obes Rev* 2007; **9**: 409–419.

112. Gregersen NT, Flint A, Bitz C, Blundell JE, Raben A, Astrup A. Reproducibility and power of ad libitum energy intake assessed by repeated single meals. *Am J Clin Nutr* 2008; **87**: 1277–1281.

113. Ordovas JM. Genotype-Phenotype associations: modulation by diet and obesity. *Obesity* 2008; **16**(Suppl. 3): S40–S46.

114. Milner JA. Nutrition in the ‘omics’ era. *Forum Nutr* 2007; **60**: 1–24.

115. EU Framework 6 Diogenes Project. <http://www.diogenes-eu.org/> (accessed 14 January 2010).