Methodological issues in the assessment of satiety

By M. Barbara E. Livingstone, Paula J. Robson, Robert W. Welch, Amy A. Burns, Martin S. Burrows and Caroline McCormack

ABSTRACT

Satiety is notoriously difficult to assess because of the considerable overlap between physiological and cognitive factors in its development. Short-term studies of satiety are typically based on a variation of the classic preload paradigm while medium-term studies involve observations of food intake, where some or all of the foods may be covertly manipulated. However, both short- and medium-term studies have generated highly variable outcomes, depending on the exact methodology used. Methodological issues that need to be considered when designing and interpreting satiety studies include 1) the use of free-living or laboratory studies, 2) the sensitivity and statistical power of the study, 3) subject selection, 4) antecedent diet of the subjects, 5) the formulation of the preload, 6) the use of subjective ratings of satiety, 7) the time interval between preload and subsequent test meal(s), 8) the formulation of the test meal(s) and 9) use of ad libitum vs fixed diet regimens in medium-term studies.

Keywords: Appetite, food intake, hunger, satiation, satiety

Introduction

Research interest in the regulation of human food intake, and the role that satiety plays in it, has been active since the 1960s. However, at present, the overwhelming amount of literature in this area presents a confused and confusing picture, largely due to procedural differences between studies and over-interpretation of the outcomes of individual studies. The aim of this review, therefore, is to evaluate some of the key methodological issues and aspects of experimental design which may, unwittingly, have exacerbated the problem.

Satiety – what is being measured?

It is widely believed that foods differ in their satiating power or efficiency and that this may be due in part, to their nutritional composition. The concept of satiating efficiency may be defined as the capacity of a consumed food to suppress hunger and decrease subsequent food intake (1-3). However, for reasons of precision and clarity, Blundell 1979 (4) has proposed that a distinction should be made between two separate but overlapping processes which determine the satiating efficiency of foods: satiation and satiety. Satiation, sometimes called short-term or intra-meal satiety (5) refers to the events during the course of an eating event which bring eating to an end. It is usually assessed by the volume or weight of food eaten and its energy and macronutrient composition. On the other hand, satiety (post-ingestive or inter-meal satiety (5)) is defined as the suppression of further intake after eating has ended. It may be assessed in terms of its intensity (the amount of food consumed at a subsequent eating event) and strength (the duration of the suppression of hunger). Taken together, the mediating processes (sensory, cognitive, post-ingestive, post-absorptive) involved in satiation and satiety are often referred to as the satiety cascade (Figure 1) (6). Hence, the satiety cascade provides the conceptual framework for experiments on the satiating effects of foods. However, because physiologically-derived eating cues are so inextricably linked with cognitive, learned cues, it is debatable whether the former can be dissociated and tested using short-term experimental paradigms.

Study designs

Studies of the short-term regulation of food intake are typically based on the preload paradigm which was first developed in the 1960s. Usually these studies are carried out within part or all of a single day. Subjects are presented with precisely prepared food(s), matched for taste, appearance and other cognitive properties, but varying in energy and/or macronutrient composition. The research question being posed 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) (7). After a variable time delay,
Methodological issues in the assessment of satiety

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 (hunger, desire to eat, fullness etc) 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 volunteers in a diary. Depending on the volume and composition of the preload and the time lapse before the test meal challenge, the experiment attempts to analyse the respective roles of post-ingestive/pre-absorptive and post-absorptive mechanisms in the regulation of food intake.

In medium-term studies, volunteers 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. 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.

The main outcome variables assessed in both short- and medium-term studies are total energy and macronutrient intakes, and sometimes also energy balance. However, studies of appetite regulation are notoriously difficult to conduct because of the considerable overlap between physiological and cognitive factors in the development of satiety. Hence the assessments made are potentially sensitive to many details of the experimental design.

Nevertheless, the apparent simplicity of the preload experimental design, coupled with the high degree of control that can be exercised in laboratory based trials, has led to a plethora of preload studies and generated a literature which is complex, often contradictory and open to every conceivable interpretation. Some general conclusions may be drawn from this research. Firstly, there is a general tendency to compensate, at least partially, for differences in the energy, but not macronutrient, content of covertly manipulated meals or preloads. Secondly, there is a wide individual variation in the efficiency of the compensatory response. Thirdly, subjective hunger ratings broadly mirror the effects of the preloads on food intake. Finally, the general, but by no means total, consensus from short-term studies supports the notion of a hierarchy in the satiating efficiency of the macronutrients. Protein is the most potent appetite suppressant, followed by carbohydrate and then fat, although the position of fat at the bottom of the hierarchy remains controversial (8–17). It is important to emphasise, however, that these general conclusions apply to mixed diets, but not necessarily to the pure macronutrients. Moreover, the source of protein (18-20), the type of fat (21,22) or form of carbohydrate (23-27) may influence intakes in the short-term, although their long-term significance has not been evaluated. It is also inconceivable that nutrients will exert a consistent effect on satiety because of moderation by a range of intricate and overlapping dietary and non-dietary factors.

Therefore while simple in rationale, conclusions derived from preloading studies must be based on a careful evaluation of the specific experimental conditions used. 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.

Methodological issues

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. Causes or mechanisms can only be clarified and internal validity ensured if measurement of eating behaviour is as accurate and precise as possible. In this context, tightly controlled laboratory studies offer the highest degree of sensitivity and control over potentially confounding variables and provide the optimum conditions for disentangling the determinants of eating behaviour. However, even when subjects are naïve to the purpose of the experiment, the notion that it is possible to fully separate the cognitive and physiological dimensions of eating behaviour under controlled conditions, is unlikely.

On the other hand, to satisfy the demands of external validity, the extent to which the outcomes of laboratory studies can be extrapolated to free-living conditions needs to be established. One of the major problems with short-term laboratory studies is that they are often deliberately designed to minimise learning about post-ingestive effects of eating which would be expected to be highly meaningful over periods of longer experience. The arguments in favour of a more naturalistic approach to the study of eating are obvious (28) and clearly it is vital to validate the findings of laboratory studies in more realistic settings. In practice this is extremely difficult because, of necessity, this is likely to involve measurements of habitual food intake which are prone to bias, particularly towards under-reporting of energy (29,30) and differential mis-reporting of the macronutrients (31,32). 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.

The issue of external validity will always be a concern for laboratory focused studies on appetite. It is essential, therefore, 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 usual eating behaviour and circumstances (33-36).

The sensitivity and statistical power of the study design

Many appetite studies, particularly those using the preload paradigm, have failed to resolve meaningful differences between experimental treatments simply because of insufficient power. Thus, while an effect size of 0.05% may be statistically significant in epidemiological studies, 10% is a more appropriate effect size in appetite studies (37).

Negative results may be attributed to a number of factors. Firstly, the absolute energy content or differences in macronutrient composition of the preload may not be of sufficient magnitude to allow detection by physiological mechanisms. Secondly, the duration of the interval between preload and test meal (time course of the preloading) may have been too long to allow the detection of otherwise, significant effects. Thirdly, the study sample may have been too small. A sample size of less than 20 is not uncommon in studies of this type.

In order to account for large inter-subject variability as well as to increase statistical power, a within-subject crossover study design is advocated in appetite research (7). By allowing subjects to serve as their own controls, studies may be more sensitive to individual variation. Nevertheless, a within-subject study design is not without its own potential problems. In particular, repeated exposures to alternative treatments could facilitate a learning component (38). This could be overcome by making more than one observation for each subject for each treatment condition, although in reality this is probably un-
likely because of practical and financial constraints. Finally there is always the risk of fatigue effects, although allowing ample time between study sessions should help to minimise monotony and boredom effects.

**Subject selection**

A recognised problem in extrapolating the results of appetite studies to any possible wider implications is that subjects are often selected on grounds of convenience rather than representativeness. Subjects who volunteer for such studies may be more likely to have specific expectations, beliefs and attitudes about food which could undermine any physiological appetite signals. For example, older subjects have been found to eat much less at lunchtime than young adult males (39,40). This is possibly due to perceptions of what are acceptable amounts of free food to eat on the part of the former, while the latter may be responding more opportunistically.

Given the diversity of subject variables that could confound experimental results, all subjects need to be routinely screened at the stage of recruitment to allow subjects to be excluded, or grouped according to common characteristics. Key characteristics include age, gender, socio-economic status, body weight, adiposity, history of overweight, current dieting status, dietary restraint and disinhibition, psychopathology, exercise habits, eating attitudes, smoking and stage in the menstrual cycle.

**Subject beliefs or knowledge about manipulations**

Most preload studies have used covert experimental manipulations in order to control for the influence of cognitive cues on subsequent food intakes. However, controlling for these cues and their possible physiological repercussions is extremely difficult. Even if food is administered in a blind fashion, orosensory factors may not be fully masked.

Moreover, when subjects are observed in several experimental conditions, they are more likely to learn quickly what is expected of them, and thus may be more susceptible to the demand characteristics of the experiment. It is conceivable, therefore, that prior knowledge, beliefs or expectations about the test foods and their energy or macronutrient contents may affect responses to experimental manipulations. In laboratory studies which have particularly focused on these issues, there is evidence that manipulations of information about the energy and nutrient contents of food does influence subsequent food intake (41,42) and subjective ratings of hunger and fullness (43,44). These observations highlight the difficulty in dissociating beliefs and perceptions of food from physiological satiety signals.

When this is the objective, in so far as it is possible, subjects should be unable to detect through orosensory or any other cognitive cues, the energy and/or macronutrient content of what they are eating.

**Antecedent diet of the subjects**

Antecedent levels of energy depletion and physical activity are potentially important confounders in appetite studies. However, failure to monitor and/or standardise them is common, making it difficult to interpret differences both within and between studies. Control of antecedent diet will be particularly important in sub-groups 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 pre-requisite, it is vital that physical activity, fasting period and alcohol intakes are standardised prior to testing in order to ensure compatibility in glycerogen stores.

**Preload formulations**

Undoubtedly, one of the reasons why short-term studies have generated highly variable outcomes is due to differences in the size and composition of the preloads. Discrepancies in the absolute energy content, macronutrient composition, state (solid vs. liquid), weight or volume, (8,16,45,46,48), sensory (16,18,19,49-51) and cognitive (52,53) characteristics of preloads could all potentially influence the outcomes of studies.

The energy loads of the manipulations appear to be particularly critical. Hence, the relatively small energy differences of preloads within studies (13,54) may have been responsible for yielding negative results with respect to energy compensation capacities. For example, it is likely that the controversy over the putative role of sweeteners and sweetness in appetite control may have been largely attributable to the small magnitude of the experimental manipulations employed (55). The satiating effects of the macronutrients has also been shown to vary according to the energy content of manipulated foods. It is only at intermediate (>1.65 MJ) or higher (>3.30 MJ) energy loads that the accepted order of satiety emerges, with protein at the top and fat at the bottom (56). This is also compatible with the observation that protein, relative to the other macronutrients, appears to be particularly satiating only above a critical threshold level of intake (9-13,51,57). However, a confounding factor in these studies was the form of the preload used, since the greater satiating effect of protein was mainly observed using solid preloads (normally familiar foods) (9,11,57), but not with liquid stimuli (13,51). In addition, given that protein exerts a potent effect on satiety, it follows that levels should be kept constant when comparing the relative satiating properties of carbohydrate and fat, otherwise it could confound any potential effects (35).

The preload paradigm, inadvertently, may have helped to fuel the controversy about the position of fat at the bottom of the satiating hierarchy (58). To gain a complete understanding of the effects of fat on appetite, it is imperative to consider its action not only on satiety but also on satiation, necessitating studies of satiating efficiency and compensatory responses (17).

In conclusion, preload formulations should always be dictated by the research issue being addressed. Pre-testing should be done to ensure that the manipulated foods are appropriate in terms of composition, weight, volume and other sensory characteristics. All covariates should be controlled for in covert manipulations such that any differences in the effects of the stimuli can be attributed solely to the post-ingestive physiological responses. Due appreciation should be paid to the fact that eating is as much a function of the time of day and habit, as it is of satiety. Consequently, the time of day at which the preload is offered and the appropriateness of the food for that time of day need to be considered. Whenever possible, double blind conditions should be observed, and control conditions should always be ensured, either by use of a no preload or a placebo treatment.

**Subjective ratings of satiety**

In order to assess the physiological and psychological dimensions of appetite sensations, fixed point (category) scales and visual analogue scales (VAS) are widely used, particularly the latter. Typically the VAS procedure uses 100 or 150 mm horizontal lines anchored at each end with the extremes of the subjective feeling to be quantified, e.g., “not at all hungry” (0 mm) and “as hungry as I have ever felt” (100 mm) in the case of the assessment of hunger. Subjects are instructed to rate the sensation being experienced according to how they define the line. Multiple measures are taken at repeated time intervals, ranging from as little as 5 minutes to over 60 minutes. Quantification of the measurement is done by measuring the distance from the left end of the line to the mark. Traditionally these scales have been constructed on paper (59) but electronic methods (60,61) offer many advantages.

The main advantages of VAS and category scaling are their ease of design, administration and data handling but there
are a number of theoretical problems associated with their use (62). In the case of category scaling it is impossible to calibrate highly subjective experiences such as palatability along a continuum of equal intervals. Similarly, it cannot be assumed that VAS are measuring the absolute intensity of a sensation. Thus, it cannot be inferred that a mark of 40 mm along a VAS for hunger rating indicates that the intensity of hunger is half that of a rating of 80 mm. Nevertheless, given the sensitivity to small changes in ratings, they should be able to detect changes in the direction or magnitude of a particular sensation. Another criticism of the VAS is the reluctance of subjects to make full use of the scale, preferring either to avoid extreme responses or to record only these responses.

The question of whether VAS ratings provide valid and reproducible indices of appetite sensations is frequently raised, but is difficult to resolve since interpretation of subjective responses is highly dependent on the subject population, experimental manipulations and statistical treatment of the results. Good reproducibility, particularly within subjects has been observed using correlation or paired rank sum analysis (63-65), but less consistent results have been noted in studies (37,61,66) which have applied the more appropriate statistical procedure of the coefficient of repeatability (65). Flint et al. 2000 (37) have concluded that despite large repeatability coefficients, VAS are reliable for single meal protocols, but that in order to avoid type 2 errors, careful attention should be paid to the measurement parameters of interest, sensitivity and power calculations.

An objective assessment of the validity of VAS is even more problematic. In the short-term, validity may be determined by calculating the extent to which subjective ratings are correlated with subsequent food intake or predict changes in food intake in response to dietary manipulations. Under controlled and free-living conditions, self ratings of hunger and appetite, desire to eat and prospective food consumption are correlated with short-term food intakes (33,63,65,68-73). Other studies have failed to demonstrate such a relationship (8,74-76) which implies that there are physiological, social, and methodological circumstances where the relationship may be weakened or lost. However, it is highly likely that there may be a methodological basis for the conclusions drawn, since the way in which the correlation coefficients are calculated can have a profound effect on the magnitude of the correlations and hence, on the conclusions drawn (77).

By definition, VAS ratings, by their subjective nature, are difficult to quantify, interpret and compare between subjects and such data must never be accepted uncritically. Nevertheless, when analysed and interpreted appropriately, they can reveal important information about the processes controlling eating behaviour. This is so, whether they are correlated with food intake, or indeed, dissociated from it.

**Interval between preload and test meal**

The major purpose of preload studies is to assess the extent to which physiological mechanisms can compensate for the ingestion of a preload at the subsequent meal. Multiple physiological mechanisms are invoked at varying times during the post-ingestive / pre-absorptive phase and post-absorptive phases of satiety. Therefore, the duration of the interval between the preload and the subsequent test meal will be decisive in determining the extent of subsequent energy and/or macronutrient compensation (15). If the purpose is to challenge the effect of orosensory and gastrointestinal factors on satiety the delay should be <30 minutes. If, on the other hand, post-absorptive inhibitory effects are being investigated, the delay needs to be much longer, but not so long that the effects of the preload have decayed to the point where they are no longer detectable.

There is considerable variation between studies in the time lapse between preload and the subsequent meal. However, many study protocols do not justify the length of this interval. This has ranged from no time delay (78,79) to several hours (8,51,57). In general, the more proximal (20-30 mins) the two events, the better the accuracy in energy compensation (22,53,80). The efficiency of the compensation is less good as the time lapse between the preload and the test meal increases (1-5 hours) (4,13,15,51,81,82).

Unfortunately, many study protocols designed to assess the relative satiating properties of the macronutrients have failed to account for the fact that the time course of the post-absorptive satiating effects of each of the macronutrients (and the form in which they are eaten) is highly variable. Therefore, by definition, the optimal time for observing these effects is also variable. For example de Graaf et al. (51), have commented that the 4 hour interval between preload and the subsequent meal in their study may have masked differences in the satiating properties of the macronutrients which have been observed after 2 hours (8,57). Undoubtedly, differences in the interval between preload and subsequent meal could account for much of the variability in the results of preload studies, and this highlights the need for a more standardised approach on this key issue. At the very least, all study protocols should be able to justify the time interval in relation to the research question being addressed. If not, decisions based largely on arbitrary criteria will merely add to the confusion.

**Formulation of subsequent meal**

In preload studies the most important criteria for the subsequent meal(s) is that it/they should be sensitive to the experimental manipulations of the preload and the direction (increased or decreased) 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 (83), they are no substitute for monitoring of test meal intakes under tightly controlled 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. Depending on the purpose of the study, the foods may be of variable nutrient composition (42,53) or of fixed nutrient composition (10). Pilot testing should establish that the foods are acceptable to the subjects.

However, despite the logic behind it, offering buffet style meals could be counterproductive. Since it is at variance with the usual eating style of the majority of people it will not necessarily guarantee a sensitive experimental protocol (84). Therefore, while an element of choice is clearly desirable, the range of foods should not be so varied that it risks undermining physiological satiety signals (85). As an alternative to the buffet style meal, a menu from which subjects pre-select in advance from a range of foods, may have merit.

**Ad libitum vs fixed diet regimens in medium-term studies**

The equivocal results of medium-term experiments on the effects of manipulation of dietary composition illustrate how they are potentially sensitive to details of experimental design. 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 lib access to their usual foods, outside of the experimental meals (86). Thus, when all food items have been manipulated, with few exceptions (87,88), the obligatory shift in the energy density of the diet causes corresponding changes in energy intake, resulting in little or no com-
pensation in subsequent intake (36,89-33). In contrast, in the majority of studies where free choice has been accommodated alongside partial manipulation of the diet in the experimental design, complete and immediate compensation is generally observed (35, 94-96). 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 in practice this would be difficult to achieve. All of these methodological issues clearly need to be taken on board when interpreting the results of the literature on medium-term studies and their implications to the free-living situation.

**Conclusion**

Currently, the major emphasis in the literature on appetite control is on the results of acute laboratory based studies. Unfortunately, while the contribution of these studies has been vital in promoting a better understanding of the differential effects of nutrients on satiety and energy intake, over-generalisation of the outcomes has been one of their pitfalls. Further progress in unravelling the interrelationships between hunger, satiety and nutrient intake would be best served by a more rigorous standardisation of definitions and procedures in this area of research. At the very least, much confusion and controversy could be avoided if interpretation of the results of individual studies was confined to the specific research question posed and in the context of the specific experimental conditions used.

**REFERENCES**

Methodological issues in the assessment of satiety


