Food Science and Technology Letters
ISSN: 0976-982X & E-ISSN: 0976-9838, Volume 5, Issue 1, 2014, pp.-052-058.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000214

COMPARISON OF EXPERIMENTAL DESIGNS AND MEASUREMENT TYPES: TWO SATIETY STUDIES

DAI H.[1]*, ANDON M.[2] AND HERSKOVIC J.E.[2]

[1]Research Development and Clinical Investigation, Children's Mercy Hospitals and Clinics, Missouri, United States.
[1]Department of Pediatrics, University of Missouri-Kansas City, United States.
[1]Department of Informatic Medicine and Personalized Health, University of Missouri-Kansas City, United States.
[2]Center for Research Quality and Innovation, ConAgra Foods Inc., Nebraska, United States.
*Corresponding Author: Email- hdai@cmh.edu

Received: July 31, 2013; Accepted: April 24, 2014

Abstract- Satiety is subjective, variable and difficult to measure. We compared the utility of a dichotomous versus a continuous measure of subjective satiety and a crossover versus a stratified parallel design. Study participants consumed a control or test meal and recorded appetite sensations for the next 5 hours. Significant treatment and time effects were found using the dichotomous measure. Only a time effect was observed for the longitudinal continuous measure. The cross-over design yielded inconsistent results and a significant treatment by day interaction. Our study suggests that discrete measures, with less psychological/rating variability, could be more powerful and robust than continuous measures in assessment of small to moderate satiety effects. We advocate incorporating both discrete and continuous measures in measure of complex sensation. For studies with confounding adaptation effects, a stratified parallel design might help reduce the experimental halo when the within-subject variation is larger than the between-subject variation.

This study provides a direct comparison of four satiety measures after food consumption and two commonly used designs. The findings may help practitioners design experiments, select response options, and determine measurement scales based on study objectives, the potential impact of confounding factors and the expected degree of difference in satiety effects. The results about variable selection and experimental comparisons might be extended to other disciplines in case of (subjective) measurements of human behavior, motivation and perception.

Keywords- Satiety, Dichotomous Measure, Continuous Measure, Likert Scale, Crossover Design, Parallel Design

Citation: Dai H., Andon M. and Herskovic J.E. (2014) Comparison of Experimental Designs and Measurement Types: Two Satiety Studies. Food Science and Technology Letters, ISSN: 0976-982X & E-ISSN: 0976-9838, Volume 5, Issue 1, pp.-052-058.

Introduction

Foods with greater ability to produce fullness or limit hunger may help control energy intake during both weight loss and weight maintenance. Even small downward shifts in energy intake, if maintained over long periods of time, could have enormous benefits in societies with a high prevalence of overweight and obesity [1,2]. It is of great value to gauge perception of food intake and how long food can deliver fullness and suppress hunger. In clinical studies, researchers have employed satiety studies to assess dietary supplement and compounds that can help patients control food intake and lose body weight in a short time period [3,4].

Satiety, similar to other human behaviors, sensations, and motivations is subjective. Although the physical effect of satiety can be inferred by measures such as weight loss and blood glucose, the subjects' perception of hunger, fullness and desire for food is primarily based on subjective measures. Instruments used to measure the perception of satiety are often continuous in nature such as visual analogue scales or Likert scales [5-7]. Discrete or dichotomous measures of satiety are employed less frequently perhaps because they are believed to produce more qualitative or categorical, rather than quantitative information. Similarly, randomized cross-over experimental designs, rather than parallel group designs, are often employed within studies of subjective satiety because they offer the benefit of each participant serving as their own control [6,8,9].

The comparisons between continuous measures and dichotomous measures have been widely conducted on a theoretical and inferential basis. The power and sample size for both measures can be calculated at a certain level of pre-determined effect size and type I error rate. However, there are few experiments that provide a head to head comparison for these two types of measures in satiety or other disciplines. Some practitioners have an impression that continuous scales are able to describe and characterize complex phenomena by measuring their exact degree or level. As a result, continuous measures are more discriminating than dichotomous measures while dichotomous measures might be too simple to catch the information in a complex and variable situation.

Are continuous measures always better to measure satiety? Previous studies have identified some issues with continuous measures: 1) the continuous assessment is often influenced by the assessors' past experience, memory, social occasion, environmental factors, and experimental manipulations. Some work has found that the visual analogue scale (VAS) is not reproducible in relation to repeated protocols [10]; and, 2) due to different perceptions, a wide range of response systems can lead to high variability. In a study with small treatment differences, signals might be overlooked because of noise.

In this study, we compare four different approaches to measure satiety. The aims of this study are to: 1) compare variability and statistical power between continuous and dichotomous response options to satiety questions; 2) compare crossover design to parallel design with stratification to the baseline hunger; 3) compare onetime measures versus longitudinal measures of satiety outcomes; and 4) compare the measures in different designs to assess reliability and reproducibility of these methodologies. The interaction between the confounding factors and the rating systems is investigated, along with the discussion on how to select an appropriate design and questionnaire to improve the signal to noise ratio.

Materials and Methods

Participants

Participants provided written informed consent before participation and were randomly selected from a database in Omaha, NE (n>10000). To avoid the adaptation effect between different satiety experiments, only subjects who were the first time participants to satiety studies were recruited. Men and women were tested in separate sessions. The participants met the following inclusion criteria: BMI between 25 and 30; age 25 through 55 years; regular and normal Dutch eating habits (three main meals including breakfast) as assessed by a questionnaire on health and lifestyle; willing to comply with the study procedures; and willing to eat the foods provided. Participants were excluded if they reported food allergy or sensitivity (wheat, milk, eggs, nuts, etc.); unexplained weight loss or weight gain of > 2kg in the month prior to pre-study screening; practice sports or exercising >10 hours a week; slimming or medically prescribed diet; vegan, vegetarian, or macrobiotic lifestyle; being pregnant or lactating; current smoking; regular medications (except vitamins and minerals); chronic or acute illness; alcohol abuse or alcoholism; or routinely drinking caffeinated beverages throughout the day (caffeinated beverages in the morning only was acceptable).

Experimental Design

Separate participant groups were tested in cross-over (n = 21) and parallel group (n = 56) experiments. The number of subjects per group was comparable in these two experiments. In the cross-over design, participants were required to finish their usual breakfast at home by 8 am. For breakfast, food items and calories were not specified but needed to be kept consistent for the two separate testing days which were separated by a one-day wash out period. Up to 500 ml of water was allowed between 8 am and 10 am. Subjects checked in to the testing facility at 11:00 am and reported food intake during the morning. Baseline measures were collected at 11:45 am, a test meal was served at 12 pm, and subjects remained at the testing facility for the next 5 hours during which they recorded the measures of subjective satiety every 30 minutes. The sensory and satiety test timetable is listed in [Table-1].

For the parallel group design, foods comprising a reduced calorie diet were provided for 2 days prior to the satiety test. Subjects checked-in at the testing facility at 8 am, were fed a standardized breakfast, and stayed at the testing site for the remainder of the day. Baseline measures were collected at 11:45 am and used to stratify and randomize participants into treatment groups. Every two subjects with the closest baseline hunger scores were randomly assigned to each of the two test meals. Test meals were fed at 12 pm and measures of satiety recorded every 30 minutes for the 5 hours post-meal period. In both experiments, no beverage of any kind was included with the test meal and participants were required to finish the entire meal within 20 minutes. Participants stayed in a large multipurpose room at the testing facility with access to non-food related movies, magazines and music. Conversations among subjects were prohibited. A moderator supervised the activity and was available to answer questions. During the 5 hour post-meal period, up to 1 liter of water was permitted. Pre-tasting diet is a confounding factor that might influence the satiety ratings. A controlled calorie diet can help align baseline satiety status. By comparing the results with and without this experiential change, we can assess the degree of influence from this confounding factor and also evaluate the satiety measures' sensitivity and robustness.

Table 1- Sensory and Satiety Test Timetable

Time	Activity
6:00 a.m.	Breakfast
8:00 a.m.	Fast
11:00 a.m.	Arrival
11:45 a.m.	Baseline of Continuous and Dichotomous Satiety Measure
12:00 p.m.	Meal
12:20 p.m.	Immediate Response of Satiety and Sensory
12:30 p.m.	Continuous and Dichotomous Satiety Measure 1
1:00 p.m.	Continuous and Dichotomous Satiety Measure 2
1:30 p.m.	Continuous and Dichotomous Satiety Measure 3
2:00 p.m.	Continuous and Dichotomous Satiety Measure 4
2:30 p.m.	Continuous and Dichotomous Satiety Measure 5
3:00 p.m.	Continuous and Dichotomous Satiety Measure 6
3:30 p.m.	Continuous and Dichotomous Satiety Measure 7
4:00 p.m.	Continuous and Dichotomous Satiety Measure 8
4:30 p.m.	Continuous and Dichotomous Satiety Measure 9
5:00 p.m.	Continuous and Dichotomous Satiety Measure 10, Stop Watch

Measurements

Satiety was measured using Continuous and Discrete Measurements through the questionnaire provided in [Table-2]. Continuous measures include the immediate response right after meal, the 11 point hunger and fullness Likert scale, and a stopwatch measure. The discrete measure was the question "if offered your favorite food or dessert I would" with the possible responses being "eat it" or "decline it". A summary of measurements is provided in [Table-3].

Test Meals

Isoenergetic meals with two different macronutrient profiles were tested. There was no significant difference between the meals in appearance, texture, flavor, the amount of food, aftertaste, and overall liking as measured by standard 9-point hedonic scales. Neither the participants, nor the study personnel conducting the satiety test were aware of the compositional difference between the meals.

Statistical Methods

Demographics and Immediate Responses were analyzed by analysis of variance (ANOVA). The discrete measure of satiety was analyzed by logistic regression. For Hunger Likert Scores, a generalized mixed model was fitted to the data. The inter-subject and intrasubject correlations were analyzed by compound symmetric variance-covariance matrix. The Likert scales were also treated as discrete variables and analyzed by non-parametric and logistic proportional odds regression models. These analyses generated consistent results with analysis from generalized mixed model. Another common approach of analyzing Likert scales and other continuous measures is to dichotomize them using certain cutoffs. We have also attempted the dichotomization but did not see any improvement (results not shown).

We investigated the effect and interactions of age, gender, treatment and time. Factors such as body mass index and baseline hunger that might influence the study outcomes were considered as covariates in the analyses. In general, covariates will summarize variance and improve statistical power; however, because adding covariates will reduce degree of freedom, non significant covariates that account for little variance were removed from model. The carry over effect and correlations among subjects have been modeled using general mixed model for both fixed effects and random effects.

The satisfaction about the amount of food and time to feel hunger were compared among treatments (control vs. test). For Hunger Likert Scores and Dichotomous Measures, we compared 1) the treatments at each time point; 2) the intercept and slope of satiety curve over time; 3) changes at each time from baseline to determine how long hunger was significantly suppressed, and 4) the area under satiety curves. For consistent findings from different methods, we selected the most representative results for publication. Statistical differences are stated at 95% confidence level ($P<0.05$).

Power analysis indicated that the study would have a power level of 0.80 to detect a moderate level of satiety effect in suppressing hunger and/or delivering fullness after meal. The power analysis took into account multi-dimensional measures of satiety profiles, including hunger, fullness, thirst, desire for snack, desire for dessert along with time to feel hunger recorded by stopwatch. A larger power level would be reached if the treatment effect were persistent or increasing over time. All the covariates such as gender, race, BMI etc have been considered and adjusted in the modeling. The carry over effect and correlations among subjects have been modeled using general mixed model for both fixed effects and random effects.

Table 2- Sensory and Satiety Questionnaire

	Question												Scale	Ask Time
Sensory Questionnaire	How much do you like this entrée OVERALL?												Dislike Extremely Dislike Very Much Dislike Moderately Dislike Slightly Neither Like or Dislike Like Slightly Like Moderately Like Very Much Like Extremely	Right after meal 12:20PM
	How much do you like the OVERALL APPEARANCE of this entree?													
	How much do you like the OVERALL FLAVOR of the entree?													
	How much do like the OVERALL TEXTURE of the entree?													
	How much do you like the AFTERTASTE of the entrée													
	How satisfied are you with the AMOUNT OF FOOD offered by the entree?													
	Right now, if offered my favorite food or dessert I would												eat it decline it	
Satiety Questionnaire	Not at all hungry	0	1	2	3	4	5	6	7	8	9	10	Extremely hungry	11:45 (Baseline); Every half of an hour from 12:30PM-5PM
	Not at all full	0	1	2	3	4	5	6	7	8	9	10	Extremely full	
	Not at all thirsty	0	1	2	3	4	5	6	7	8	9	10	Extremely thirsty	
	Not at all interested in a snack	0	1	2	3	4	5	6	7	8	9	10	Extremely interested in a snack	
	My mood is extremely good	0	1	2	3	4	5	6	7	8	9	10	My mood is extremely bad	
	Time to first noticed hunger												Stopwatch	Anytime from 12:30PM to 5PM
	Time to "so hungry have to eat"													

Table 3- A Summary of Measurements

	Measurements	Description
Continuous	Immediate Response	Right after meal, subjects evaluated satisfaction about the amount of food along with overall liking, appearance liking, flavor liking, texture liking, and aftertaste on the standard 9-point hedonic scales
	Hunger Likert Score	Hunger, fullness, thirst, desire for a snack, thirst and general mood were evaluated at baseline and every 30 minutes for up to 5 hours post meal using 11-point Likert scale.
	Stopwatch	Subjects used stopwatch to record the time to first notice hunger and time to feel so hungry that they need to eat
Discrete	Dichotomous Measure	Subjects were asked whether they would decline their favorite food or dessert at baseline and every 30 minutes for up to 5 hours post meal.

Results

Results from Cross-over Design

Nine females and twelve males with BMI between 25 and 30 consumed a Control meal and a Test meal in two days separated by one-day wash out period. The dichotomous measure showed time had a linear effect in the logistic model and the likelihood of food rejection decreased at a rate of 0.33±0.08, ($P<0.0001$). We observed an adaptation effect as subjects in Day 2 had higher food rejection percentages (odds ratio=1.7, 95% confidence interval [1.1, 2.6], $P=0.0187$). The results indicate that the subjects in Control group were more likely to turn down their favorite food or dessert for five hours post meal (odds ratio=2.3, 95% confidence interval [1.5, 3.6], $P=0.0002$) [Fig-1].

Degrees of satisfaction about the amount of food from the immediate response were not significantly different between the two treatments (Control=6.2±0.4, Test=6.3±0.4, $P=0.8574$). According to the stopwatch measure, the Control sample delivered fullness feeling for 3.0±0.3 hours before the subjects first noticed hunger but the length of time was not significantly longer than for the Test sample (2.7±0.3, $P=0.3846$). “Time to need to eat” by the stopwatch measure was not analyzed due to 81% missing data.

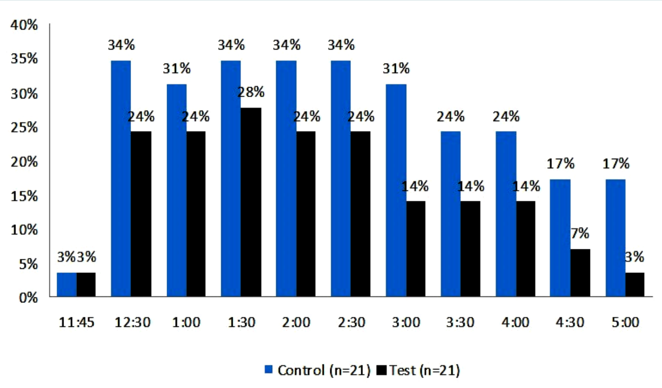

Fig. 1- Dichotomous Measure: Percentage of subjects who would decline their favorite food or dessert in Crossover Design (Treatment Effect: p=0.0002, Time Effect: p<0.0001, Adaptation Effect: p=0.0187)

In the Crossover Design, subjects came in two days to taste both meals. There was a treatment by day interaction at baseline. As a result of this undesirable interaction, the Hunger Likert Score was analyzed by taking the difference from baseline. Time had a linear effect (*P*<0.0001) with the Hunger Likert Score increasing by 0.90±0.06/hour. A significant adaptation effect in the assessment of satiety was found as subjects gave significantly higher Hunger Likert scores (difference between Day 1 and Day 2=1.6±0.2, *P*<0.0001). However, there was no significant treatment effect between Test and Control in the Hunger Likert Score (P=0.2801) [Fig-2].

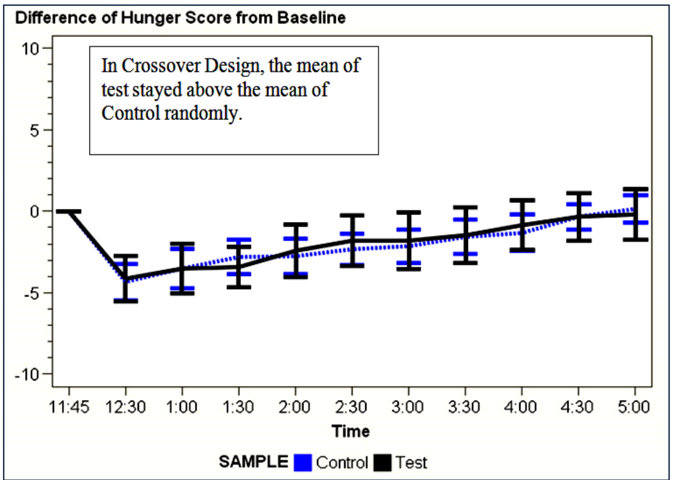

Fig. 2- Continuous Measure of Adjusted Hunger Curve from Crossover Design (Treatment: p=0.2801, Time: p<0.0001, Adaptation: p=0.0001)

Results from Parallel Design

For the dichotomous measure time had a linear effect in the logistic model and the likelihood of food rejection decreased at a rate of 0.52±0.06 (P<0.0001). The results indicate that subjects in the Control condition were more likely to turn down their favorite food or dessert as compared to the Test condition (odds ratio=2.3, 95% confidence interval [1.5, 3.4], P=0.0002) [Fig-3].

The experiment was repeated with better control of baseline diet and the adaptation effect. Fifty-six subjects were randomly assigned to one of the treatment groups (27 in Control, and 29 in Test; 28 females and 28 males). The results from experiment #2 confirmed the findings from the first experiment. No significant differences in the satisfaction about the amount of food (immediate response) and time to feel hunger (stopwatch); although, the Control mean trended in a more favorable direction [Table-4].

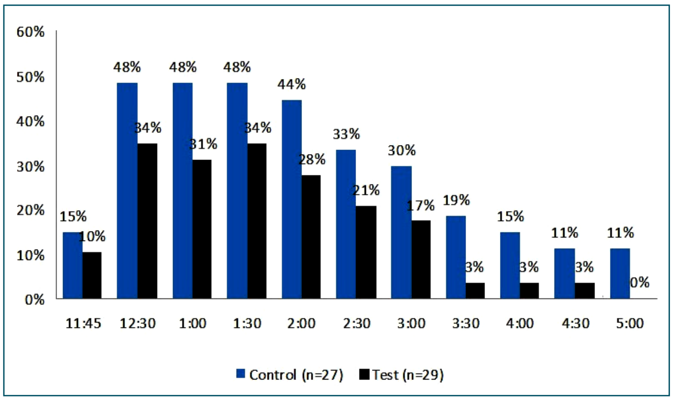

Fig. 3- Dichotomous Measure: Percentage of subjects who would decline their favorite food or dessert in stratified parallel design (Treatment Effect: p= 0.0002, Time Effect: p<0.0001, Adaptation Effect: Avoided by Design)

Due to the pre-testing diet control, parallel design and stratification, the adaptation effect was removed and all the repeated measures were found to be at parity at the baseline measurement. As a result, no baseline adjustment was needed for the analysis of the satiety scores. Time had a significant linear effect (*P*<0.0001): the Hunger Likert Score increased by 0.93±0.03/hour. There were no significant differences in the rates of change between the satiety profiles from experiment 1 to experiment 2 (P>0.05), indicating the satiety experiments are repeatable. In experiment 1, the mean Hunger Likert curves from the Control and Test samples crossed over at different time points. Due to the design improvement, in experiment #2, the Control trended to be superior to the Test with the parallel curves. Despite the more rigorous design, no significant treatment effect could be demonstrated between Test and Control in Hunger, Fullness and Desire for a Snack (*P*=0.4268 for Hunger Likert Score, 0.1415 for Fullness Likert and 0.0650 for Snack Likert Score) [Fig-4],[Fig-5].

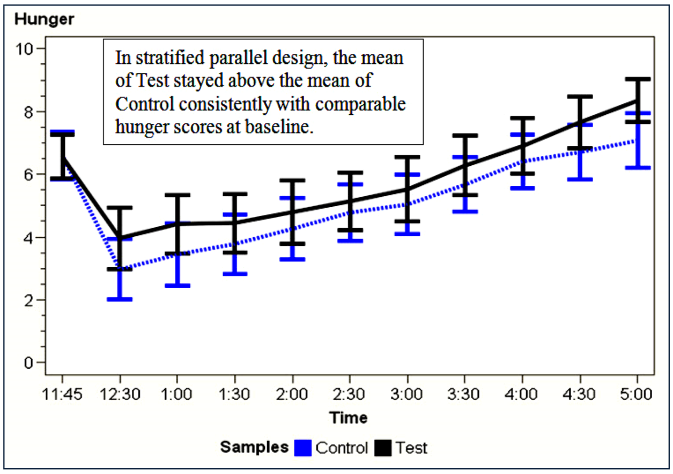

Fig. 4- Continuous Measure of Hunger Curve from stratified parallel design (Treatment: p=0.4268, Time: p<0.0001)

Table 4- One Time Measure of Sensory and Satiety Right after the Meal[1]

		Cross-over Design			Parallel Design		
		Control	Test	P	Control	Test	P
	Overall Liking	7.8±0.3	7.4±0.3	0.1593	6.6±0.4	6.4±0.4	0.9745
	Appearance Liking	7.4±0.2	7.2±0.2	0.3368	6.7±0.3	6.7±0.3	0.8662
Sensory[1]	Flavor Liking	7.7±0.3	7.5±0.3	0.4279	6.6±0.4	6.0±0.4	0.5365
	Texture Liking	7.8±0.3	7.3±0.3	0.1103	6.3±0.3	6.2±0.3	0.7817
	Aftertaste	7.3±0.3	6.9±0.3	0.1557	6.5±0.4	5.8±0.3	0.2924
	Amount of Food[3]	6.2±0.4	6.3±0.4	0.8574	6.3±0.4	5.8±0.4	0.412
Satiety	first notice hunger (h)	3.0±0.3	2.7±0.3	0.3846	3±0.3	2.7±0.2	0.7028
	have to eat[4] (h)	N/A[5]	N/A	N/A	4.7±0.1	4.7±0.1	0.8319

[1]Values are least squares means ± standard deviation
[2]Sensory was measured at 12:15PM from 1=dislike extremely to 9=like extremely
[3]Amount of food was measured at 12:15PM from 1=unsatisfied extremely to 9=satisfied extremely
[4]Post meal time to “first noticed hunger” and time to “feel so hungry that I have to eat” was measured by stopwatch.
[5]Data is not presented in this table due to substantial missing values. There are no treatment effects for the existing data and no treatment effects using survival analysis that treats missing as censoring.

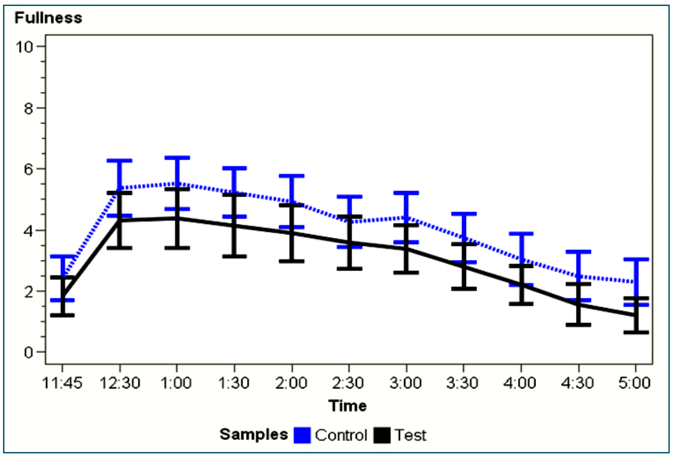

Fig. 5- Continuous Measure of Fullness Curve from Stratified Parallel Design (Treatment Effect: p=0.1450, Time Effect: p<0.0001)

Discussion

In many experiments, only Likert scale and other continuous measures that reflect the degree of sensation is designed into the experiment while simple dichotomous measures are left out of the investigation. This study explored the impacts of design and measures to three types of effects: the treatment effect, the time effect and the adaptation effect. Without loss of generosity, the treatment effect represents effects with small effect sizes, i.e. small signal to noise ratios. This type of effect is variable and typically requires larger sample sizes. Within 5 hours after a small meal, study participants were very likely to become hungary, so the time effect represents the type of effects with moderate to large effect sizes. The adaptation effect compares the difference between the first visit and the second visit in the crossover design, which represents an undesirable halo influence to study outcomes. The findings of this study can help practitioners select an appropriate design and questionnaire to extract signal from noises for satiety studies and subjective measures in other disciplines.

Dichotomous Measure vs. Continuous Measure

Our work provides a direct comparison between continuous and dichotomous measures of satiety and demonstrates that the dichotomous measure could be more powerful and robust when the effect size between two treatment groups is small, as we have seen in the treatment effect. When the effect size gets larger, as we have seen in the time effect, both continuous and dichotomous measures are able to demonstrate the difference.

It is well known that the satiety and other subjectively measured human behaviors are complex and variable. From a statistical point of view, two aspects need to be factored into consideration in design of the psychometric response system. On one hand, a rating system needs to be discriminatory by providing sufficient response options to generate spread in data. On the other hand, too many options or an oversized range may potentially inflate measurement variability and other founding factors’ impact on ratings.

Continuous measures are commonly used to capture the characteristics of human sensation. This work raises some issues that could occur when using continuous measures. The measurement has inherent variability. The wider the range in continuous measures, the more likely it could suffer from other sources of variation such as metabolism, eating habit, physical activity and baseline diet etc. In our satiety studies, the continuous measure does a good job identifying the strong signal in the time effect. However, this measure was not able to detect the weak signal in the treatment effect, especially when the measurement variability was confounded with the adaptation effect at the current sample sizes [Fig-3],[Fig-6].

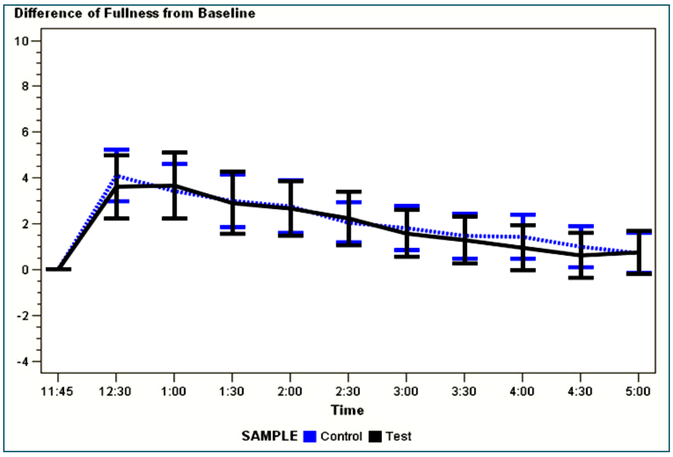

Fig. 6- Continuous Measure of Adjusted Fullness Curve from Crossover Design (Treatment: p=0.2902, Time: p<0.0001, Adaptation: p=0.0259)

A longitudinal dichotomous measure of food rejection is the only measure in this study that was able to detect the treatment effect. In both experiments, the baseline status was comparable between Test and Control. As a result, we can compare the trend of response curve that peaked after meal consumption and gradually decreased over time [Fig-1],[Fig-3]. When the longitudinal continuous measure was noisy along with other sources of variation, experimenters could mistakenly attribute all sources of variation to a single trait and infer no satiety effect. Randomization and recruitment control did not generate comparable baseline hunger scores.

Dichotomous measures are advantageous because satiety measures are subjective and the psychological effect is evident, but the pattern and pathway of impact cannot be separated from the treatment effect. Blundell [11] argues that hunger, appetite and satiety are hypothetical constructs that we use to conceptualize our perception of sensation or motivations which are themselves indirectly linked to process which influence our behavior. Psychological reaction, physiological control, social occasion and environmental background can all be reflected in hunger ratings and thus inflate the variation of satiety ratings. A simple dichotomous question will help align the perception of hunger by generating a hunger threshold. Once the hunger intensity exceeds the threshold, an individual will express a desire for food. As compared with the continuous variables, the dichotomous measure has less between-subject and within-subject variability. As a result, subjects are able to differentiate the products' satiety effect by accurately and consistently responding to the dichotomous question.

There have been concerns that a dichotomous measure of satiety (yes/no) is not very relevant in most studies, because researchers are generally interested in the evolution of appetite/satiety over time. Our experiments shows that both continuous and dichotomous measures in longitudinal studies are able to validly display the satiety curve over time and a significant time effect was found in both methods from both experiments. The dichotomous measure exhibited a similar satiety profile, which peaked at 12:30 and gradually decreases over time, as did the continuous Hunger Measure. [Fig-1] to [Fig-6]. For the dichotomous measure a percentage can be calculated and used to describe the likelihood of food rejection. By pooling the responses from different time points together we are able to perform trend analysis, thus increasing the power [Fig-1], [Fig-3]. Furthermore, the profile of dichotomous measure is highly correlated with the profile of the 9-point Fullness measure (R-square=0.93 in [Fig-7]).

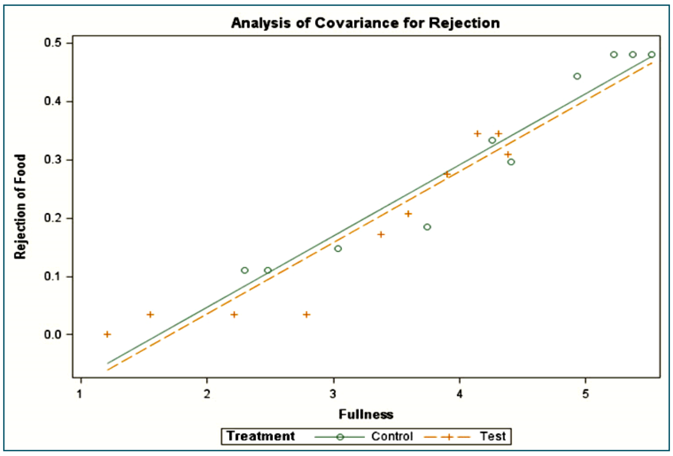

Fig. 7- Correlation between Continuous Fullness Measure and Dichotomous Measure in stratified parallel design (R-square=0.93)

Crossover Design vs. Parallel Design/Stratification

The secondary objective in this work is to compare the Crossover (within subject) design versus the Parallel (between subject) design. Crossover experimental designs are commonly employed in the experiments comparing two or more treatments as they offer the advantage of participants serving as their controls. However, the underlying assumption of a crossover is that within subject variability is smaller than the between subject variability. Furthermore, the order of treatment is assumed to have equal carryover effects and no residual effects. These assumptions were not met within the context of our study. This may indicate that participants, although instructed to do so, did not in fact prepare equally for the two different test days or reflect the complex nature of measuring subjective satiety. In either case, even after adjusting for the treatment by day interaction by expressing the responses as change from a baseline, no consistent pattern of change in hunger overtime was observed between the treatments. In fact, the continuous measure of hunger for the two treatments crisscrossed multiple times during the 5 hour post-meal time period [Fig-2] and subjects started higher on the hunger scale for Day 2 as compared to Day1 (reaching significance for the group that got Control-Test as a sequence). In contrast, employing a parallel group design with stratification for baseline hunger resulted in a consistent, albeit non-significant, pattern of difference over time with the treatment by time lines running parallel ([Fig-3] versus [Fig-4], [Fig-5] versus [Fig-6]).

Central Location Test

In both experiments, we chose to use a central location testing (CLT) methodology. Stubbs et al. [10] has pointed out that the reliability and validity of the subjective measures of hunger and satiety appear to be more pronounced under controlled laboratory conditions relative to real life. Many social and environmental elements can interfere with the perception of hunger. People might have different hunger profiles due to different activity intensities during work or leisure time. The activity after meal may also have an interaction with the physiological control of appetite. It is hard to isolate the objective (unconditioned or physiological) components from the subjective (conditioned or perceived) components in term of the causality to the sensation of hunger. Thus, the extent to which the difference between the test meals observed under these controlled conditions translates to a free-living condition is unknown.

One Time Measures vs. Repeated Measure

The degree of Satisfaction about the amount of food was asked after each meal. The advantage of this method is the easy administration and direct correlation with the sensory questionnaire. Similar to the sensory questions, this question used a standard 9 point hedonic scale and it can be incorporated into the sensory ballot and administered to subjects at the same time to explore the association between satiety and sensory. The limitation of this satiety measure is that this onetime measure is unable to catch a dynamic change in human behavior.

Time to first notice hunger and the time to have to eat are one time variables that are measured using a stopwatch. While the degree of satisfaction about the amount of food focuses on the impact of food right after a meal, time to feel hunger focuses on the time interval to the critical change in satiety status. Therefore, this one time variable can be linked with other repeated measured satiety variables for a better understanding of the change in satiety over time.

Conclusions

Often the dichotomous measures are not given full consideration in complex experiments and crossover design is often used to control the experimental noise. This study quantified the impact of study design and different types of subjective measures to three types of effects: meal treatment, time, and adaptation. A significant degree of adaptation occurred with the cross-over design. This issue was avoided with the stratified parallel group design. Both designs and all measures were adequate to detect the predictable decrease in hunger after consuming a meal and the gradual increase in hunger over a 5-hour post-meal time period. However, only the longitudinal dichotomous measure of food rejection was able to detect the meal treatment effect. Our study demonstrates that dichotomous measures of satiety coupled with a stratified parallel group design can be more robust than the more typically employed cross-over design with continuous measures.

It is not our intent to devalue the Likert scale other continuous measures or the crossover design. Instead, we demonstrate the strength of dichotomous measures and parallel design in certain scenarios. For products with small to moderate satiety effect sizes, the longitudinal dichotomous measure of responses to their favorite food is shown to be the most acute and robust. As compared to the dichotomous measure, the continuous measures offer wider response ranges to provide the discrimination of different traits. However, the relatively complex rating system could generate more variations from the system itself and from the confounding factors. For treatments that have very pronounced and prolonged satiety effects, repeated continuous measures are able to efficiently summarize and analyze the data over time but the efficacy of the continuous measures reduces as the satiety effect gets smaller. We advocate incorporating both discrete and continuous measures in measure of complex sensation. Due to the adaptation effect, stratified parallel design is more beneficial in showing the satiety trend. Parallel Design in conjunction with stratification on baseline hunger appear to provide a better control of noise and to depict the satiety curves for the continuous repeated measures of hunger, fullness and desire for snack. Compared to the one-time measures, the repeated measures allow for trend analysis to provide a stronger power level by reducing standard error of estimated variables and demonstrate the change of effect size over time. The same recommendations can also be extended to pain studies, educational evaluations and other psychometric measures.

Acknowledgements

The authors thank Ursula Lowczyk, Casey Smith and Shawn Ward for their dedication and expertise in conducting the in-life portion of these experiments.

Conflicts of Interest: None declared.

References

[1] Dall T.M., Fulgoni V.L., Zhang Y., Reimers K.R., Packard P.T. & Astwood J.D. (2009) *Am. J. Health Promot.*, 23, 412-422.

[2] Dall T.M., Fulgoni V.L., Zhang Y., Reimers K.R., Packard P.T., Astwood J.D. (2009) *Am. J. Health Promot.*, 23, 423-430.

[3] Reid M. & Hetherington M. (1997) *Neuro & Biobehav Reviews*, 21(3), 295-308.

[4] Halton T.L. & Hu F.B. (2004) *J. Am. Coll. Nutr.*, 23(5), 373-385.

[5] Solah V.A., Kerr D.A., Adikara C.D., Meng X., Binns C.W., Zhu D.A. and Prince R.L. (2010) *Appetite*, 54, 485-491.

[6] Hursel R., Van Der Zee L. & Westerterp-Plantega M.S. (2010) *Br. J. Nutr.*, 103, 775-780.

[7] Chaput J.P., Gilbert J.A., Gregersen N.T., Pedersen S.D. & Sjodin A.M. (2010) *Appetite*, 54, 583-586.

[8] Ratliff J., Leite J.O., De Ogburn R., Puglisi M.J., Vanheest J. & Fernandez M.L. (2010) *Nutr. Resr.*, 30, 96-103.

[9] Schroeder N., Gallaher D.D., Arndt E.A. & Marquart L. (2009) Appetite, 53, 363-369.

[10] Stubbs R.J., Hughes D.A., Johnstone A.M., Rowley E., Reid C., Elia M., Stratton R., Delargy H., King N. & Blundell J.E. (2000) *Brit J. Nutr.,* 84, 405-415.

[11] Blundell J.E. (1979) *Nutrition and Lifestyles*, Applied Science Publishers, London, 21-42.