



In search of the most reproducible neural vulnerability factors that predict future weight gain: analyses of data from six prospective studies

Sonja Yokum, ¹ Ashley N. Gearhardt, ² and Eric Stice³

¹Oregon Research Institute, Eugene, OR 97403, USA, ²Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA, and ³Department of Psychiatry, Stanford University, Stanford, CA 94305, USA

Correspondence should be addressed to Sonja Yokum, Oregon Research Institute, 1776 Millrace, Eugene, OR 97403, USA. E-mail: sonjas@ori.org.

Abstract

We tested if we could replicate the main effect relations of elevated striatum and lateral orbitofrontal cortex (OFC) response to high-calorie food stimuli to weight gain reported in past papers in six prospective datasets that used similar functional MRI (fMRI) paradigms. Participants in Study 1 ($N = 37$; M (mean) age = 15.5), Study 2 ($N = 160$; M age = 15.3), Study 3 ($N = 130$; M age = 15.0), Study 4 ($N = 175$; M age = 14.3), Study 5 ($N = 45$; M age = 20.8) and Study 6 ($N = 49$; M age = 31.1) completed fMRI scans at the baseline and had their body mass index (BMI) and body fat (Studies 4 and 6 only) measured at the baseline and over follow-ups. Elevated striatal response to palatable food images predicted BMI gain in Studies 1 and 6 and body fat gain in Study 6. Lateral OFC activation did not predict weight gain in any of the six studies. The result provide limited support for the hypothesis that elevated reward region responsivity to palatable food images predicts weight gain. Factors that make replication difficult are discussed and potential solutions considered.

Key words: striatum; OFC; prospective; weight gain; cross-replication

Several studies have tested the hypothesis that individuals who show elevated reward region responsivity to high-calorie food cues and images are at elevated risk for future weight gain, which is a cornerstone to the dynamic vulnerability model of obesity (Stice and Yokum, 2016). Greater responsivity of the lateral orbitofrontal cortex (OFC) to cues that predict the impending presentation of the images of high-calorie foods predicted weight gain in adolescents (Yokum *et al.*, 2011). Elevated responsivity in the nucleus accumbens (NAcc) to high-calorie food images predicted weight gain in adults (Demos *et al.*, 2012). Greater caudate response to commercials for high-calorie foods predicted weight gain in adolescents (Yokum *et al.*, 2014). Further, elevated lateral OFC response to cues that signal

impending receipt of tastes of high-calorie chocolate milkshake predicted weight gain in adolescents; this effect replicated in split-halves of the sample (Stice *et al.*, 2015). It is reassuring that two of these studies found peaks in the striatum and two found peaks in the lateral OFC, which are two regions that are involved in reward valuation. However, two other studies did not produce evidence that elevated responsivity of these regions or other reward valuation regions to high-calorie food cues or images showed main effect relations to weight gain in adolescents (Stice *et al.*, 2010; Stice and Yokum, 2018). Thus, published studies provide somewhat mixed support for the hypothesis that elevated reward region responsivity to high-calorie food cues and images predicts future weight gain. Some of the inconsistencies

Received: 2 December 2019; Revised: 4 January 2021; Accepted: 29 January 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

regarding neural vulnerability factors that predict weight gain may have resulted because the studies used different functional MRI (fMRI) paradigms and examined samples that varied in age and weight range. However, it would have been more reassuring if the findings showed greater convergence. Further, the small sample sizes used in many of these studies likely contributed to the mixed findings because this increases the risk for both false-positive and false-negative findings.

Given the limited evidence of reproducibility regarding neural vulnerability factors that predict weight gain, we thought it prudent to conduct a cross-replication study with multiple datasets to determine whether the results reproduce. Specifically, we tested whether we could replicate previous findings showing that elevated reward region responsivity to high-calorie food cues and images predicts weight gain (Demos et al., 2012; Stice et al., 2015; Yokum et al., 2011, 2014) in six datasets, using a priori small volume correction (SVC) analysis within the striatum and lateral OFC. We focused on new datasets that used similar fMRI paradigms to those used in past studies, which include exposure to high-calorie food images (Demos et al., 2012), high-calorie fast-food commercials (Yokum et al., 2014) and cues signaling impending delivery of a palatable high-calorie food taste (Stice et al., 2015). We evaluated the reproducibility of these relations. Further, to increase the sensitivity to detect small effects, we combined data across studies that used similar paradigms to examine the effects of brain reward response on weight gain across samples.

Subjects and methods

Participants

Study 1. Participants were 44 adolescent girls recruited from Eugene, Oregon. Data from seven participants were excluded due to excessive head movement during the scan (i.e. within-run movement exceeded 2 mm in translational movement and 2° in rotational movement), resulting in a final sample of 37 participants [M age = 15.5 ± 1.0; M baseline body mass index (BMI) = 24.2 ± 5.0; 5.6% American Indian/Alaska Native, 2.9% African-American, 82.9% European-American and 8.6% mixed racial heritage]. More detailed descriptions of the sample and measures are discussed in a previous publication (Stice et al., 2010).

Study 2. Participants were 162 lean adolescents recruited from Eugene, Oregon. fMRI data of two participants were collected with an acquisition error. These participants were excluded from the analyses, resulting in a final sample of 160 participants (80 females; M age = 15.3 ± 1.1; M BMI = 20.8 ± 1.9; 11.5% Hispanic-American, 0.6% American Indian/Alaska Native, 0.6% Asian-American, 67.9% European-American and 19.4% mixed racial heritage). More detailed descriptions of the sample and measures are discussed in Stice and colleagues (2015) which present the results of whole brain analyses testing the relation between neural response to receipt and anticipated receipt of palatable food and body fat gain.

Study 3. Participants were 135 lean adolescents recruited from Portland, Oregon. fMRI data of five participants were excluded from the analyses ($n = 3$ acquisition error and $n = 2$ incomplete data), resulting in a final sample of 130 participants (70 females; M age = 15.0 ± 0.9; M BMI = 21.2 ± 2.2; 9.2% Hispanic-American, 1.5% American Indian/Alaska Native, 6.2% Asian-American,

11.5% African-American and 71.6% European-American). In a previous publication (Stice and Yokum, 2018), we provide more details on the study and discuss main effects of left NAcc and left caudate response to palatable food images and taste of four different milkshakes on BMI gain.

Study 4. Participants were 193 adolescents recruited from Southeast Michigan. In total, 186 adolescents completed the fMRI scan. Nine participants showed excessive movement during the scan, and fMRI data of two participants were collected with an acquisition error. These participants were excluded from the analyses. This yielded 175 participants (90 females; M age = 14.3 ± 1.0; M baseline BMI = 24.1 ± 5.4; 9.1% Hispanic-American, 1.7% American Indian/Alaska Native, 1.1% Asian-American, 13.1% African-American, 61.3% European-American, 0.6% other, 8.0% mixed and 5.1% unknown) for neural analyses. See Gearhardt and colleagues (Gearhardt et al., 2020) for more details on the sample and measures.

Study 5. Participants were 48 overweight and obese young adult women drawn from a prospective study in Eugene, Oregon. Three participants showed excessive head movement during the scan and were therefore excluded from the analyses, resulting in a final sample of 45 participants (M age = 20.8 ± 1.3; M BMI = 28.4 ± 2.9; 2.2% Hispanic-American, 4.4% American Indian/Alaska Native, 8.9% Asian-American, 75.6% European-American, 8.9% mixed racial heritage). See Yokum and colleagues (2015) for more details on the sample and measures.

Study 6. Participants were 53 overweight/obese adults recruited from Eugene, Oregon and drawn from a larger ongoing study evaluating the efficacy of a multifaceted food response and attention training intervention on weight and body fat loss. At the time of data analyses, 49 participants had completed the fMRI scan and 6-month follow-up data (M age = 31.0 ± 7.3; M BMI = 30.7 ± 3.7; 75.5% female; 17.0% Hispanic-American, 2.0% American Indian/Alaska Native, 4.1% Asian-American, 2.0% African-American, 60.6% European-American and 14.3% mixed racial heritage). All participants' data met the movement inclusion criteria (see Study 1 criteria).

In Studies 1–4, participants provided assent and parents provided written informed consent. In Studies 5–6, participants provided written informed consent. The University of Michigan Institutional Review Board approved Study 4. The Oregon Research Institute Institutional Review Board approved all other studies.

In Study 4, exclusion criteria were a BMI percentile of <5%, lifetime psychiatric disorder (including eating disorders), current use of psychotropic medications or illicit drugs, or fMRI contra-indicators (e.g. presence of metal implants). In all other studies, individuals were excluded when they reported binge eating or compensatory behavior in the past 3 months, current use of psychotropic medications or illicit drugs or psychiatric disorder, dairy allergies, or fMRI contra-indicators (see Supplementary Material for more information).

Measures

Body mass index. BMI (weight in kg/height in m²) was used as a proxy measure for adiposity and was available in all studies. Height was measured to the nearest millimeter. Weight was assessed to the nearest 0.1 kg after removal of shoes

and coats. See Table 2 for BMI data points per study. BMI correlates with direct measures of total body fat such as dual energy X-ray absorptiometry ($r=0.80$ to 0.90) and with health measures including blood pressure, adverse lipoprotein profiles, atherosclerotic lesions, serum insulin levels and diabetes mellitus (Dietz and Robinson, 1998; Steinberger et al., 2006).

Percentage body fat. In Studies 4 and 6, we collected percent body fat (see Table 2). Body fat percentage was estimated with bioelectric impedance (InBody Body Composition scale) in Study 4 and with air displacement plethysmography (BodPod) in Study 6. Both measures show test–retest reliability ($r=0.92$ – 0.99) and correlate with dual X-ray absorptiometry and hydrostatic weighing estimates ($r=0.91$ – 0.99 ; Fields et al., 2002; Von Hurst et al., 2016).

fMRI paradigms

In Studies 1 and 3, participants completed a food picture paradigm and a food receipt paradigm. In Study 4, participants completed a food commercial paradigm. Participants in Studies 2 and 5 completed a food receipt paradigm. Participants in Study 6 completed a food picture paradigm. In Studies 2–4 and Study 6, we collected hunger ratings prior to the scan. See Table 1 and Supplementary Material for more details on the fMRI paradigms and hunger ratings. The main effects of the food picture paradigms and food receipt paradigms in Studies 1, 2, 5 and 6 are reported in Supplementary Material (Supplementary results and Supplementary Tables S1–S5). Main effects of the food picture paradigm and food receipt paradigm in Study 3 (Stice et al., 2013; Stice and Yokum, 2018) and of the food commercial paradigm in Study 4 (Gearhardt et al., 2020) are reported in previous publications.

Analysis

Change in BMI and body fat percentage. In Study 4, change in BMI and change in body fat percentage were assessed by calculating the difference between baseline and 1-year follow-up BMI and body fat percentage, respectively. In all other studies, change in BMI and body fat were estimated with mixed effects growth models using SAS PROC MIXED (Version 9.3). The models specified fixed and random effects

$$Y_{ij} = \beta_{0i} + \beta_{1i} \text{Time}_{ij} + r_{ij},$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i},$$

$$\beta_{1i} = \gamma_{10} + \mu_{1i},$$

Y_{ij} is the BMI or body fat score if the i th person at the j th time, β_{0i} is the intercept (defined at the baseline assessment) for the i th person, β_{1i} is the slope for the i th person and r_{ij} is the residual variance for the i th person at the j th time. The intercept has a mean fixed effect for each person, γ_{00} , and person-specific deviation around the mean intercept, μ_{0i} . Similarly, the slope has a mean fixed effect for each person, γ_{10} , and person-specific deviation around the mean slope, μ_{1i} .

These latter analyses offer a sensitive technique for modeling change in continuous variables and use maximum likelihood estimation to accommodate missing data (Singer, 1998).

fMRI data analysis. Detailed descriptions of the fMRI data acquisition and data preprocessing are provided in Supplementary Material. Neuroimaging data were preprocessed and analyzed using SPM12 (Wellcome Department of Cognitive Neurology; <http://www.fil.ion.ucl.ac.uk/spm>). At the subject level, blood oxygen level-dependent (BOLD) signal signal was modeled in a fixed effects analysis with separate regressors modeling each condition of interest during the picture, cue and commercial presentation period (Table 1). In Studies 1 and 3, to identify brain regions activated by appetizing food images in the food picture paradigm, we contrasted BOLD signal during appetizing food images vs unappetizing food images and vs glasses of water. In Study 6, we contrasted activations during appetizing high-calorie food images vs appetizing low-calorie food images and vs glasses of water. To identify brain regions activated by unhealthy fast-food commercials (Study 4), we contrasted BOLD signal during unhealthy fast-food commercials vs healthier fast-food commercials and vs phone commercials. Activation in response to the cue signaling impending receipt of the milkshake in the food receipt paradigms (Studies 1, 2, 3 and 5) was assessed by contrasting BOLD signal during milkshake cue vs tasteless solution cue. In Study 3, all milkshake variants were preceded by the same image of a milkshake to not confound the neural response to receipt with expectations (Stice and Yokum, 2018). We collapsed BOLD response across cues for all four milkshakes. The contrast assessing BOLD response to milkshake taste (i.e. milkshake receipt > tasteless solution receipt) was excluded from the food receipt paradigm analyses. All data were high-pass filtered at 128 s and first-order autoregressive (AR[1]) error was used to correct for serial autocorrelations.

The individual SPM contrasts were entered into second-level regression models with BMI slopes and intercepts as covariates. Hunger prior to the scan was included as a covariate of no interest in the models for Studies 2–4, and 6. In Studies 5 and 6, we also included intervention (0 = no, 1 = yes) as a covariate of no interest as participants in these studies were randomized to either a weight loss intervention (Study 5 $n=24$; Study 6 $n=17$) or a control condition (Study 5 $n=21$; Study 6 $n=32$). In Studies 4 and 6, we used similar regression models to test if BOLD response to unhealthy fast-food commercials and appetizing high-calorie food pictures predicted body fat gain.

We performed SVC analyses within the left NAcc (Montreal Neurological Institute [MNI] coordinates: $x=-9$, $y=6$, $z=-4$; Demos et al., 2012), left caudate (MNI coordinates $x=-12$, $y=-7$, $z=22$; Yokum et al., 2014) and right lateral OFC (MNI coordinates $x=36$, $y=27$, $z=-15$; Stice et al., 2015) with activation peaks as centroids to define 6-mm diameter spheres. Peaks were considered significant at P -values <0.05 , familywise error rate (PFWE) corrected across the small volume. Exploratory analyses tested for relations of BOLD activation within the anatomical regions of interest (ROIs) (i.e. bilateral NAcc, bilateral caudate and bilateral lateral OFC) to BMI gain and body fat gain to examine if there were any other significant effects within these brain regions. We used mask images for the caudate and lateral OFC from the WFU Pickatlas (Maldjian et al., 2003). Due to the lack of a NAcc mask in the WFU Pickatlas, we used the NAcc mask from the Pauli atlas (Pauli et al., 2018). Peaks were considered significant at a peak level of $P<0.05$ FWE corrected across the total number of voxels across the ROI. We also conducted exploratory whole brain analyses to test for significant relations of brain activation outside of our *a priori* hypothesized ROIs to BMI gain/body fat gain (see Supplementary Material for methods and results).

Power to detect a medium effect size ($r=0.30$) in each of the datasets was 0.45 in Study 1, 0.97 in Study 2, 0.92 in Study 3, 0.97

Table 1. Overview of samples, fMRI tasks, time of day of scans, fasting and contrasts

Study	Sample	fMRI task	Time of day scans	Last time eaten	Contrasts
Study 1 (Stice <i>et al.</i> , 2010)	37 adolescents (females only)	Food picture	Between 11 am and 6 pm	4–6 h prior to scan	Appetizing food images > unappetizing food images. Appetizing food images > glasses of water
		Food receipt	Between 11 am and 6 pm	4–6 h prior to scan	Milkshake cue > tasteless solution cue
Study 2 (Stice <i>et al.</i> , 2015)	160 adolescents (male and female)	Food receipt	Between 11 am and 6 pm	4–6 h prior to scan	Milkshake glass > water glass
Study 3 (Stice and Yokum, 2018)	130 adolescents (male and female)	Food picture	Between 1 pm and 6 pm	4 h prior to the scan	Appetizing food images > unappetizing food images. Appetizing food images > glasses of water
		Food receipt			Milkshake cue > tasteless solution cue
Study 4 (Gearhardt <i>et al.</i> , 2020)	175 adolescents (male and female)	Food commercials	Between 10:30 am and 6 pm	1–4 h prior to scan	Unhealthy fast-food commercials > healthier fast-food commercials; unhealthy fast-food commercials > phone commercials; healthier fast-food commercials > phone commercials
Study 5 (Yokum <i>et al.</i> , 2015)	45 young adults (females only)	Food receipt	Between 10 am and 6 pm	4–6 h prior to scan	Milkshake glass > water glass
Study 6	49 adults (34 females)	Food picture	Between 9 am and 5 pm	3–4 h prior to scan	Appetizing high-calorie food > appetizing low-calorie food. Appetizing high-calorie food > glasses of water

in Study 4, 0.54 in Study 5 and 0.55 in Study 6. Power to detect a small effect size ($r = 0.10$) was 0.08 in Study 1, 0.24 in Study 2, 0.20 in Study 3, 0.26 in Study 4, 0.09 in Study 5 and 0.10 in Study 6. Because three studies (Studies 1, 5 and 6) had sample sizes that were not large enough to detect a medium effect ($r = 0.30$), we decided not to use additional Bonferroni corrections to correct for the number of brain regions and contrasts to avoid the possibility that it leads us to overlook real effects in the form of Type II errors (Lieberman and Cunningham, 2009). For all outcomes, we estimated effect sizes (r) based on the reported Z-values and sample size using a formula from Rosenthal (1991).

Post hoc analyses. We ran post-hoc exploratory regression analyses using data from multiple studies to increase sensitivity to detect small effects of brain response in our *a priori* ROIs on

BMI gain. Statistical analyses were performed using the Statistical Package for the Social Sciences 24 (SPSS software package (SPSS 24, SPSS Inc., Chicago, IL, USA). Due to the difference in fMRI paradigms, analyses were conducted separately for studies including the food receipt paradigm (Studies 1–3 and 5: $N = 363$) and those including the food picture paradigm (Studies 1, 3 and 6: $N = 216$). Given that the food commercial paradigm in Study 4 differed from the other two paradigms (e.g. video clips vs static images), we excluded this study from the analyses. For the food receipt and the food picture paradigms, we extracted the main effect parameter estimates at the individual level in the left NAcc, left caudate and right lateral OFC using MarsBar (<http://www.marsbar.sourceforge.net>). The parameter estimates were exported to SPSS. For the food picture paradigm analyses, we only included data from the contrast appetizing (high-calorie) food images > glasses of water because this was

Table 2. Overview of the average baseline BMI and average change in BMI over follow-up per Study. In Study 4, change in BMI and change in body fat were assessed by calculating the difference between baseline and 1-year follow-up BMI and body fat percentage, respectively. In all other studies, data from baseline and all follow-ups were used in random intercept, mixed effects growth curve analyses (SAS Inc. version 9.3) to model BMI change and body fat change

Study	M baseline BMI (s.d.)	M BMI gain (s.d.)	Measurements
Study 1	24.46 (5.42)	0.23 (0.43)	Baseline, 1 year, 2 years and 3 years
Study 2	20.81 (1.93)	0.51 (0.49)	Baseline, 1 year, 2 years and 3 years
Study 3	21.15 (2.25)	0.46 (0.45)	Baseline, 1 year, 2 years and 3 years
Study 4	24.10 (5.35)	0.45 (1.49)	Baseline and 1 year
Study 5	28.20 (2.85)	-0.14 (0.30)	Baseline, 3 months, 6 months, 1 year and 2 years
Study 6	30.59 (3.87)	0.01 (0.78)	Baseline, 1 month, 3 months and 6 months

the only contrast that was similar across the three studies. BMI change over follow-up was the dependent variable. Independent variables included baseline BMI, sex, age, study (dummy coded), parameter estimates from the left NAcc, left caudate and right lateral OFC. Further, we examined whether study (dummy coded) moderated the effects of neural response on BMI gain. Prior to these latter analyses, we centered the parameter estimates within the left NAcc, left caudate and right lateral OFC on their mean to maximize interpretability and minimize multicollinearity (Aiken and West, 1991). To adjust for multiple comparisons, the level of significance was set to $P < 0.013$. Power to detect small effects ($r = 0.10$) across datasets was 0.47 for the food receipt paradigm dataset ($N = 363$) and 0.31 for the food image paradigm dataset ($N = 216$). Power to detect medium effects ($r = 0.30$) was 0.99 for both datasets.

Results

Descriptive statistics

Table 2 presents the average change in BMI over follow-up per study. In Study 4, the average change in body fat percentage from baseline to 1-year follow-up was $M = -0.80 \pm 3.50$, suggesting that there was significant variation in change in body fat percentage. In Study 6, the average change in body fat percentage over 6-month follow-up was $M = -0.24 \pm 1.70$. There were no significant differences in BMI change between participants randomized to the weight loss intervention and control condition in Study 5 ($t(43) = -1.97, P = 0.06$) and Study 6 ($t(67) = 0.35, P = 0.73$).

Relations of BOLD activity to food images, fast-food commercials and food cues to BMI gain over follow-up

SVC analyses testing for relations between BOLD activity and BMI gain using the threshold of $pFWE \leq 0.05$ revealed two effects. In Study 1, BOLD activation in the left caudate [MNI coordinates = $-15, -10, 23, Z = 2.69, pFWE = 0.05, r = 0.44, 95\%$ confidence interval (95% CI): 0.13, 0.67] in response to the contrast appetizing food images > glasses of water predicted BMI gain over a 3-year follow-up (Figure 1). In Study 6, BOLD activation in the left NAcc (MNI coordinates: $-6, 5, -4, Z = 2.73, pFWE = 0.05, r = 0.39, 95\%$ CI: 0.12, 0.60) in response to the contrast appetizing high-calorie food images > glasses of water predicted BMI gain over 6-month follow-up (Figure 2A). In Studies 2–5, there were no significant effects of neural activity in these regions on BMI gain (left NAcc $M r = 0.05$, left caudate $M r = 0.05$, right lateral OFC $M r = 0.01$; see Supplementary results and Supplementary Table S6).

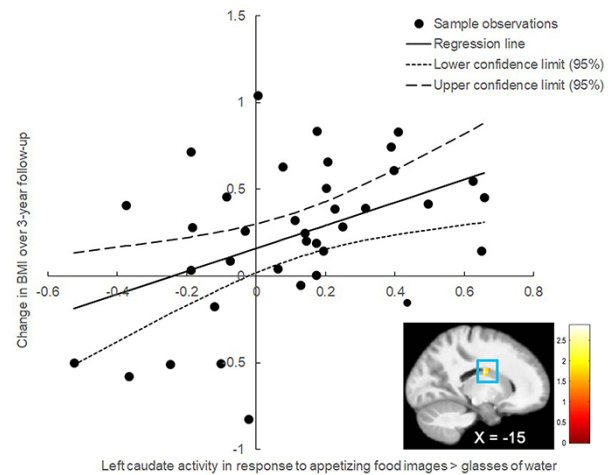


Fig. 1. In Study 1, BOLD activity in the left caudate (MNI coordinates = $-15, -10, 23, Z = 2.69, pFWE = 0.05, r = 0.44; 95\%$ CI: 0.13, 0.67) in response to the contrast appetizing food images > glasses of water predicted BMI gain over a 3-year follow-up.

Exploratory analyses tested for effects of BOLD activation within the anatomical ROIs (i.e. bilateral caudate, bilateral NACC and bilateral lateral OFC) to examine if there were other significant effects within these brain regions. There were no significant relations between BOLD activity within the anatomical ROIs and BMI gain.

Relation of neural responsivity to body fat change over follow-up

In Study 6, elevated BOLD activation in the left caudate (MNI coordinates: $-9, -4, 20, Z = 3.43, pFWE = 0.008, r = 0.49, 95\%$ CI: 0.24, 0.68) in response to the contrast appetizing high-calorie food images > glasses of water predicted body fat gain over 6-month follow-up (Figure 2B). In Study 4, there were no significant effects of BOLD activity in the NAcc ($M r = 0.09$), caudate ($M r = 0.02$) and OFC ($M r = -0.01$) on change in body fat over 1-year follow-up (see Supplementary Table S6). There were no significant correlations between BOLD activity within the anatomical ROIs and body fat gain.

Post hoc tests of the effects of BOLD response to food images and cues on BMI gain across datasets

Although there was a trend-level effect of left caudate (partial $r = 0.14, P = 0.05$) response to the contrast appetizing food

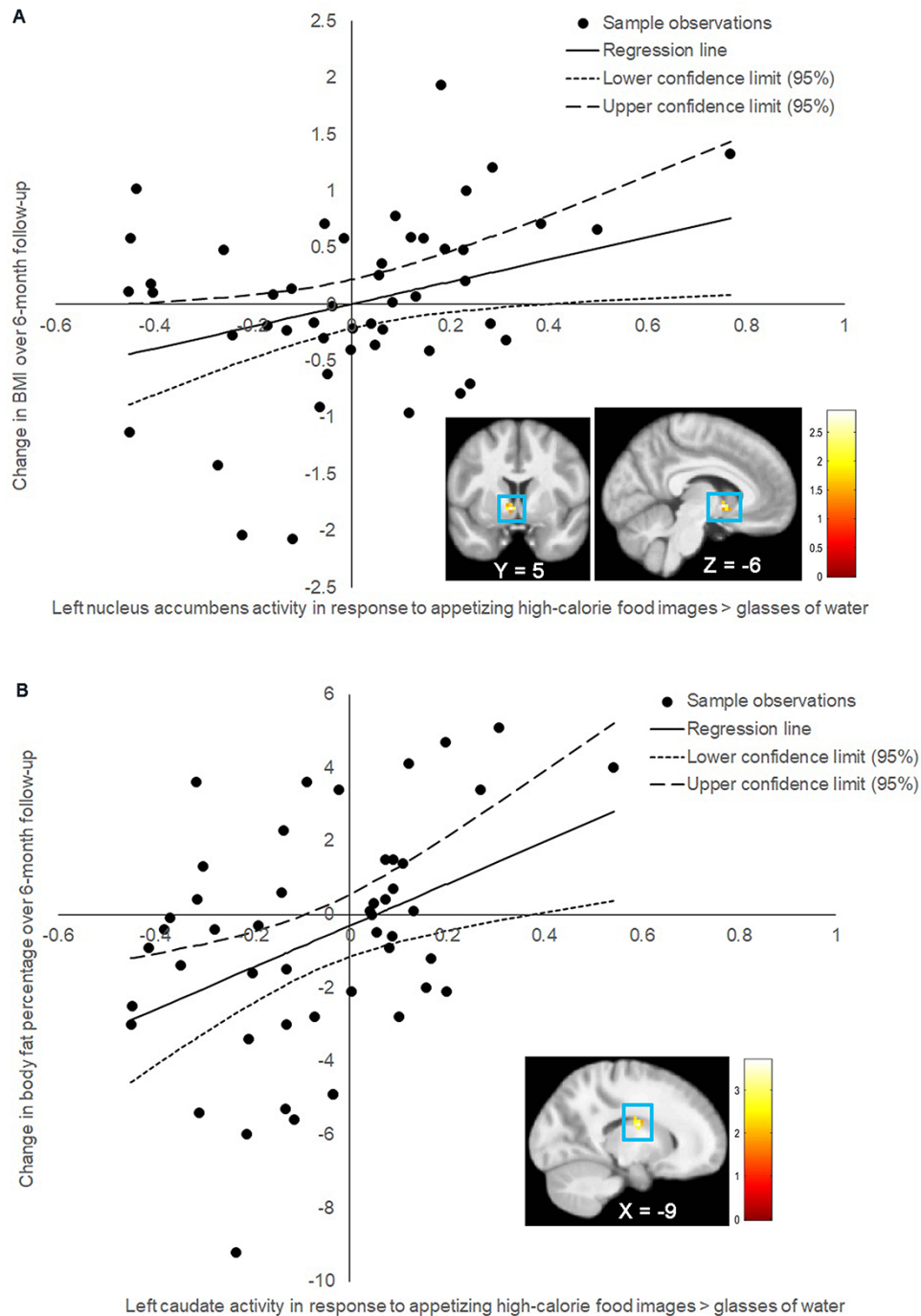


Fig. 2. In Study 5, BOLD activity in (A) the left nucleus accumbens (MNI coordinates: $-6, 5, -4, Z = 2.73$, $p_{FWE} = 0.05$, $r = 0.39$; 95% CI: 0.12, 0.60) and (B) the left caudate (MNI coordinates: $-9, -4, 20, Z = 3.43$, $p_{FWE} = 0.008$, $r = 0.49$; 95% CI: 0.24, 0.68) in response to the contrast appetizing high-calorie food images > glasses of water predicted BMI gain and body fat gain over 6-month follow-up, respectively.

images > glasses of water on BMI gain across the datasets ($N = 216$) (Table 3), this effect did not survive adjustment for multiple comparisons ($P < 0.013$). There were no significant effects of left NAcc (partial $r = 0.13$, $P = 0.07$) and right lateral OFC (partial $r = -0.06$, $P = 0.39$) response to the contrast appetizing food images > glasses of water on BMI gain (Table 3). Study 6 significantly moderated the effects of left NAcc (partial $r = 0.20$, $P = 0.005$) and right lateral OFC (partial $r = 0.18$, $P = 0.01$) response to the contrast appetizing food images > glasses

of water on BMI gain: the positive main effects of neural response in these two regions on BMI gain were significantly stronger in Study 6 compared to the other two studies (Studies 1 and 3).

There were no significant main effects of left NAcc (partial $r = -0.02$, $P = 0.76$), left caudate (partial $r = -0.01$, $P = 0.91$) and right lateral OFC (partial $r = 0.02$, $P = 0.73$) response to the contrast milkshake cue > tasteless solution cue ($N = 363$) on BMI gain (Table 3). Study did not moderate the effects of neural

Table 3. Main effects of the left NAcc, left caudate and right lateral OFC in response to the food receipt paradigm and food image paradigm on increases in BMI over follow-up across datasets, while controlling for baseline BMI, sex, age and study

Variable	Standardized B	SE	t-value	P-value	Partial r
Food receipt paradigm (N = 363): milkshake cue \geq tasteless solution cue					
Left NAcc	-0.02	0.01	-0.29	0.77	-0.02
Left caudate	-0.01	0.02	-0.13	0.89	-0.01
Right lateral OFC	0.02	0.01	0.40	0.69	0.02
Food image paradigm (N = 216): appetizing (high calorie) food images > glass of water					
Left NAcc	0.13	0.19	1.84	0.07	0.13
Left caudate	0.14	0.22	1.97	0.05	0.14
Right lateral OFC	-0.06	0.20	-0.86	0.39	-0.06

activity in these regions in response to the contrast milkshake cue > tasteless solution cue on BMI gain.

Discussion

Our goal was to investigate whether elevated reward-related neural responsivity to palatable food images and cues predicts weight gain in six prospective datasets. We also performed exploratory analyses to examine effects of brain reward response on weight gain across datasets. We found that elevated BOLD activity in the left caudate in response to appetizing food images (relative to glasses of water) predicted BMI gain in Study 1. Elevated BOLD activity in the left striatum (caudate, NAcc) in response to appetizing high-calorie food images (relative to glasses of water) predicted both BMI gain and percent body fat gain in Study 6. We found a trend-level effect of elevated left caudate response to appetizing food images (relative to glasses of water) on BMI gain across datasets. The effect sizes in Studies 1 and 6 are comparable with those found in Demos and colleagues (Demos et al., 2012; $r = 0.37$) but lower than those found in Yokum and colleagues (Yokum et al., 2014; $r = 0.57$).

The pattern of findings might be interpreted as suggesting that findings from fMRI studies show low reproducibility, in that we only observed a significant effect in two of the six datasets and a trend-level effect across three datasets (N = 216). Much greater confidence could be placed in the hypothesis that individuals who show elevated reward region response to high-calorie food images or cues are at elevated risk for future weight gain had we observed similar effects in each of the six datasets or a significant effect across datasets. It could be argued that the inconsistent pattern of findings is a function of the fact that studies differed with regard to sample composition regarding age and weight, fMRI paradigms, length of follow-up and statistical power. For instance, both studies in which we found predictive effects (Study 1 and Study 6) included overweight and obese individuals. These results suggest that a history of overeating is necessary to establish a relation between brain reward response to palatable food images on weight gain, consistent with the dynamic vulnerability model of obesity (Stice and Yokum, 2016). However, we did not find significant effects of brain reward response to unhealthy fast-food commercials or to cues predicting impending palatable food delivery in two other studies with overweight and obese individuals, including one study that had a power of 0.97 to detect a medium effect size (Study 4).

It is also important to consider the power to detect a medium effect in each of the studies because it may be unrealistic to

detect significant effects in all six samples because of the multiplicative nature of power across studies. Specifically, our power to detect a medium effect was 0.45 in Study 1, 0.97 in Study 2, 0.92 in Study 3, 0.97 in Study 4, 0.54 in Study 5 and 0.55 in Study 6. This means that we only had a probability of 0.12 for detecting medium effects in all six samples ($0.45 \times 0.97 \times 0.92 \times 0.97 \times 0.54 \times 0.55 = 0.12$). This suggests that it may be unrealistic to detect the same effect in all six datasets given the compound nature of power across samples. Indeed, even with the 0.80 recommended power, the probability of actually detecting effects in six datasets would only be 0.26 ($0.80 \times 0.80 \times 0.80 \times 0.80 \times 0.80 \times 0.80 = 0.26$). These analyses suggest that cross-replication studies with large enough samples (e.g. six studies with 0.97 power) are needed to replicate findings.

One interesting pattern in the findings is that we found some support for striatal effects on weight gain in two samples and across three samples; however, we did not find significant relations between lateral OFC activation and weight gain in even one sample. This pattern suggests that elevated striatal responsivity to food images is a more reproducible predictor of future weight gain than lateral OFC responsivity to food images or cues. These findings converge with evidence that elevated striatal responsivity to monetary reward, but not elevated OFC responsivity to monetary reward, predicted onset of substance use over 1-year follow-up (Stice et al., 2013). However, it is important to acknowledge that differences in the paradigms used across studies might have contributed to the inconsistent findings regarding OFC activity as the predictor of weight gain. Indeed, the main effects (Supplementary Tables S1–S5) suggest that there is little overlap in BOLD activity in response to the different paradigms across the studies. Future studies should evaluate how specific design parameters elicit different neural responses to food stimuli.

The pattern of findings reported herein is concerning because it is certainly possible to observe much greater reproducibility of scientific findings. For instance, the effects of a dissonance-based eating disorder prevention program on the five core outcomes have replicated in 88% of the tests in the 22 randomized controlled trials conducted by several independent research teams (Stice et al., 2017). This is in contrast to the 3% of the effects reported in the published studies that replicated in the analyses conducted with the six datasets reported herein. These 22 prevention trials likewise varied in a number of factors like the fMRI studies, such as the nature of the samples (e.g. adolescent vs adult participants and healthy weight vs overweight participants), nature of the experimental design (e.g. the types of facilitators who implemented the prevention program

and the nature of the comparison conditions), sample size and statistical power. Although it is tempting to attribute non-replication to variation in such design features, it would be far more reassuring if studies addressing the same general research question generated effects that reproduce in multiple studies conducted by independent teams.

The current research has important strengths, including the examination of six different prospective datasets, objectively measured weight and body fat, long follow-ups and datasets that varied in age, sex and BMI. However, there are also key limitations. First, three of the six studies used small samples, which increase the risk for false positives (Smeets *et al.*, 2019). Because significant effects emerged in two of the smaller studies and because we did not correct for multiple comparisons across regions and contrasts, it is possible that these results are false-positive findings. Indeed, the total number of significant SVC effects (i.e. 3) is equal to the expected number of false positives (60 SVC tests \times 0.05 = 3). However, we did find a trend-level effect of caudate response to palatable food images in data across three samples ($N = 216$), suggesting that the caudate finding is a moderately reliable effect. Future studies with large enough samples are needed to attempt to replicate the present findings. Second, the studies differed in data collection approaches (e.g. time of day of the scan, hours of fasting and MRI scanners), the duration of follow-up, the time between follow-ups and the change in BMI and body fat, which may have contributed to the lack of reproducibility across samples. For example, time of day of the scans (Byrne *et al.*, 2017) and hunger (Siep *et al.*, 2009) has been suggested to affect striatal functioning. Future research should address these limitations so that stronger inferences regarding the relation of reward region responsivity to future weight gain are possible.

In conclusion, our neuroimaging results provide only limited support for the hypothesis that elevated reward region responsivity to palatable food images predicts future weight gain. Specifically, there was evidence that elevated striatal response to palatable food images predicted weight gain, converging with the findings of two previous studies (Demos *et al.*, 2012; Yokum *et al.*, 2014). On the other hand, the striatal effects only emerged in response to palatable food images and not in response to unhealthy fast-food commercials or to cues predicting impending palatable food delivery, suggesting that the relation is not particularly robust. It will be vital to determine how to generate fMRI findings that are more likely to reproduce in studies conducted by independent teams, such as by using larger samples, more standardized procedures and paradigms with more events.

Acknowledgements

All authors read and approved the final manuscript. The authors also thank the Lewis Center for Neuroimaging at the University of Oregon for their assistance in imaging for this investigation.

Funding

This work was supported by grants from the National Institute on Diabetes and Digestive and Kidney Diseases: R01 DK080760, R01 DK112762, and R01 DK102532.

Conflict of interest

The authors report no conflict of interest with respect to the content of this paper.

Supplementary data

Supplementary data are available at SCAN online.

References

- Aiken, L.S., West, S.G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.
- Byrne, J.E.M., Hughes, M.E., Rossell, S.L., Johnson, S.L., Murray, G. (2017). Time of day differences in neural reward functioning in healthy young men. *The Journal of Neuroscience*, 37(37), 8895–900.
- Charbonnier, L., van Meer, F., Johnstone, A.M., Crabtree, D., Buosi, W., Smeets, P.A.M. (2018). Effects of hunger state on the brain responses to food cues across the life span. *NeuroImage*, 171, 246–55.
- Demos, K.E., Heatherton, T.F., Kelley, W.M. (2012). Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *Journal of Neuroscience*, 32(16), 5549–52.
- Fields, D.A., Goran, M.I., McCrory, M.A. (2002). Body-composition assessment via air-displacement plethysmography in adults and children: a review. *The American Journal of Clinical Nutrition*, 75(3), 453–67.
- Gearhardt, A.N., Yokum, S., Harris, J.L., Epstein, L.H., Lumeng, J.C. (2020). Neural response to fast food commercials in adolescents predicts intake. *The American Journal of Clinical Nutrition*, 111(3), 493–502.
- Hare, T.A., Malmaud, J., Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *Journal of Neuroscience*, 31(30), 11077–87.
- Lieberman, M.D., Cunningham, W.A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4(4), 423–8.
- Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19(3), 1233–9.
- Noonan, M.P., Kolling, N., Walton, M.E., Rushworth, M.F. (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *European Journal of Neuroscience*, 35(7), 997–1010.
- Pauli, W.M., Nili, A.N., Tyszka, J.M. (2018). A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data*, 5, 180063.
- Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research*. Newbury Park, CA: Sage.
- Siep, N., Roefs, A., Roebroek, A., Havermans, R., Bonte, M.L., Jansen, A. (2009). Hunger is the best spice: an fMRI study of the effects of attention, hunger, and calorie content on food reward processing in the amygdala and orbitofrontal cortex. *Behavioural Brain Research*, 198(1), 149–58.
- Singer, J.D. (1998). Using SAS PROC MIXED to fit multi-level models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23, 323–55.
- Smeets, P.A.M., Kroese, F.M., Evers, C., de Ridder, D.T.D. (2013). Allured or alarmed: counteractive control responses to food temptations in the brain. *Behavioural Brain Research*, 248, 41–5.
- Smeets, P.A.M., Dagher, A., Hare, T.A., Kullmann, S., van der Laan, L.N., Veldhuizen, M.G. (2019). Good practice in

- food-related neuroimaging. *The American Journal of Clinical Nutrition*, **109**(3), 491–503.
- Stice, E., Yokum, S., Bohon, C., Marti, N., Smolen, A. (2010). Reward circuitry responsivity to food predicts future increases in body mass: moderating effects of DRD2 and DRD4. *NeuroImage*, **50**, 1618–25.
- Stice, E., Yokum, S., Burger, K. (2013). Elevated reward region responsivity predicts future substance use onset but not overweight/obesity onset. *Biological Psychiatry*, **73**, 869–76.
- Stice, E., Burger, K.S., Yokum, S. (2015). Reward region responsivity predicts future weight gain and moderating effects of the Taq1A allele. *Journal of Neuroscience*, **35**(28), 10316–24.
- Stice, E., Rohde, P., Shaw, H., Gau, J. (2017). Clinician-led, peer-led, and internet-delivered dissonance-based eating disorder prevention programs: acute effectiveness of these delivery modalities. *Journal of Consulting and Clinical Psychology*, **85**(9), 883–95.
- Stice, E., Yokum, S. (2016). Gain in body fat associated with increased striatal response to palatable food cues whereas body fat stability is associated with decreased striatal response. *Journal of Neuroscience*, **36**(26), 6949–56.
- Stice, E., Yokum, S. (2018). Relation of neural response to palatable food tastes and images to future weight gain: using bootstrap sampling to examine replicability of neuroimaging findings. *NeuroImage*, **183**, 522–31.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**(1), 273–89.
- Von Hurst, P.R., Walsh, D.C.I., Conlon, C.A., Ingram, M., Kruger, R., Stonehouse, W. (2016). Validity and reliability of bioelectrical impedance analysis to estimate body fat percentage against air displacement plethysmography and dual-energy X-ray absorptiometry. *Nutrition and Dietetics*, **73**, 197–204.
- Yokum, S., Gearhardt, A., Harris, J., Brownell, K., Stice, E. (2014). Individual differences in striatum activity to food commercials predicts weight gain in adolescence. *Obesity (Silver Spring)*, **22**(12), 2544–51.
- Yokum, S., Marti, C.N., Smolen, A., Stice, E. (2015). Relation of the multilocus genetic composite reflecting high dopamine signaling capacity to future increases in BMI. *Appetite*, **87**, 38–45.