



OPEN ACCESS

Edited by:

Szilvia Anett Nagy,
University of Pécs, Hungary

Reviewed by:

Owen O'Daly,
King's College London,
United Kingdom
Luke Norman,
University of Michigan, United States
Daniel John Halls,
King's College London,
United Kingdom

***Correspondence:**

Simon Maier
simon.maier@uniklinik-freiburg.de

[†]These authors have contributed
equally to this work and share senior
authorship

Specialty section:

This article was submitted to
Neuroimaging and Stimulation,
a section of the journal
Frontiers in Psychiatry

Received: 30 May 2020

Accepted: 21 July 2020

Published: 05 August 2020

Citation:

Horster I, Nickel K, Holovics L,
Schmidt S, Endres D, Tebartz van
Elst L, Zeeck A, Maier S and Joos A
(2020) A Neglected Topic in
Neuroscience: Replicability of fMRI
Results With Specific Reference
to ANOREXIA NERVOSA.
Front. Psychiatry 11:777.
doi: 10.3389/fpsy.2020.00777

A Neglected Topic in Neuroscience: Replicability of fMRI Results With Specific Reference to ANOREXIA NERVOSA

Isabelle Horster¹, Kathrin Nickel², Lukas Holovics¹, Stefan Schmidt¹, Dominique Endres²,
Ludger Tebartz van Elst², Almut Zeeck¹, Simon Maier^{1,2†*} and Andreas Joos^{1,3†}

¹ Department of Psychosomatic Medicine and Psychotherapy, University Medical Center, University of Freiburg, Freiburg, Germany, ² Department of Psychiatry and Psychotherapy, University Medical Center, University of Freiburg, Freiburg, Germany, ³ Department of Psychosomatic Medicine and Psychotherapy, Ortenau Klinikum, Offenburg, Germany

Functional magnetic resonance imaging (fMRI) studies report impaired functional correlates of cognition and emotion in mental disorders. The validity of preexisting studies needs to be confirmed through replication studies, which there is a lack of. So far, most replication studies have been conducted on non-patients (NP) and primarily investigated cognitive and motor tasks. To fill this gap, we conducted the first fMRI replication study to investigate brain function using disease-related food stimuli in patients with anorexia nervosa (AN). Using fMRI, we investigated 31 AN patients and 27 NP for increased amygdala and reduced midcingulate activation when viewing food and non-food stimuli, as reported by the original study (11 AN, 11 NP; Joos et al., 2011). Similar to the previous study, we observed in the within group comparisons (food>non-food) a frontoinsula activation for both groups. Although in AN the recorded activation clustered more prominently and extended into the cingulate cortex. In the between-group comparisons, the increased amygdala and reduced midcingulate activation could not be replicated. Instead, AN showed a higher activation of the cingulate cortices, the pre-/postcentral gyrus and the inferior parietal lobe. Unlike in the initial study, no significant differences between NP>AN could be observed. The inconsistency of results and the non-replication of the study could have several reasons, such as high inter-individual variance of functional correlates of emotion processing, as well as intra-individual variances and the smaller group size of the initial study. These results underline the importance of replication for assessing the reliability and validity of results from fMRI research.

Keywords: replicability, anorexia nervosa, food, functional magnetic resonance imaging (fMRI), neurobiology

BACKGROUND

Anorexia nervosa (AN) usually affects young women and shows high persistence rates of around 50% (1). Furthermore, it has the highest mortality of all mental disorders (2). The etiology is largely unknown, although an interplay of genetic and environmental factors is assumed (3). The AN pathophysiology consists largely of reduced weight, fear of weight gain and a distorted body perception, as well as a cognitive preoccupation with body and food related issues. For this reason, functional magnetic resonance imaging (fMRI) studies have focused on paradigms with disease-related food and body stimuli to investigate the neuronal correlation of the disorder.

The first fMRI study in AN with visual food cues (six patients, six non-patients (NP)) described greater activation of anterior cingulate cortices (ACC), left insular, and amygdala-hippocampal regions (4). Fourteen years later, a meta-analysis across nine studies applying food cues, reported increased activation of frontocingular cortices and lower activation of the parietal brain (5). However, the design and the results differed between the included studies. Three further reviews confirmed these inconsistencies (6–8) and therefore conclusions remain questionable. None of the studies were confirmed by replication, so the reported findings should not yet be regarded as established scientific knowledge.

The necessity of replications is not only increasingly recognized in the neurosciences, but in the entire scientific community (9–12). The awareness of a general lack of data replication in science, also referred to as a “reproducibility/replicability crisis” (13–16), has emerged in particular during the last decade (17). Although it is generally recognized that the replication and reproduction of scientific claims is essential in scientific research, the deficit of replications persists (9). Furthermore, there is no general agreement on the definition or directives of replication procedures (9, 16, 18, 19). The *Committee on Reproducibility and Replicability in Science* (9) suggested the following definition: “Reproducibility is obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis. (...) Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. Two studies may be considered to have replicated if they obtain consistent results given the level of uncertainty inherent in the system under study.” Other studies in the field also refer to this definition (15, 17, 20) and this publication adheres to it, too. In addition to exact definitions, the precise description of study protocols, data, and results is of importance (21). Replication serves the validation of exploratory results and therefore the transition from exploratory data into knowledge, to generate confirmable and generalizable principles (9).

There have been some replication efforts in the field of fMRI, but the studies are largely limited to NP and to motor and cognitive tasks (15, 17, 22, 23). However, Bennett and Miller (24) strongly assume that factors influencing the ability of replication (i.e., variance) are larger in emotional paradigms and in clinical populations, including eating disorders (25). Furthermore, low sample sizes, low power, and low effect-sizes, which reduce

replicability, have been generally reported in the field of fMRI research (26–28). If replication attempts failed with sample sizes of 15–30, as a consequence of low power and low effect-sizes, this would have profound influences on planning further studies with respect to number of participants and study set-ups (29).

Against this background, the objective of the present study was to replicate for the first time an fMRI study in AN using visual food and non-food stimuli. Our aim was to replicate the original study (30) with the same research question in a larger but similar sample, using the identical study design and closely following the fMRI and analysis protocol.

In the original study (30), both AN (N=11) and NP (N=11) showed an involvement of fronto-insular and ACC areas when comparing food>non-food pictures (within-group effects) (**Figure 2A**). Comparing the two groups, AN had elevated blood oxygenation level dependent (BOLD) responses of the right amygdala and less activation in midcingulate cortices (MCC).

We assume that (1) there will be different neural correlates of the food-stimuli in AN compared to NP, uncovering disease-related responses, (2) that within-group data of food>non-food pictures will show an involvement of fronto-insular and cingulate cortices, and (3) between-group data will reveal elevated BOLD responses of the right amygdala and decreased activation in midcingulate cortices (MCC) in AN compared to NP similar to our earlier results.

In addition, we assessed emotional reactions to the stimuli by rating the images after scanning.

MATERIALS AND METHODS

This study was part of a multimodal MRI study, which assessed structural, metabolic and other functional data [see, e.g., (31–36)]. We replicated the aforementioned food paradigm with 31 AN and 27 NP.

In the following, we first describe Material and Methods of the current study and point towards differences with the earlier study in the second section.

Current Replication Study

Sample and State of Participants

For sample description see **Table 1**. All participants were studied in the second half of the menstrual cycle or the equivalent stage with estrogen and progesterone when taking oral contraception in the current investigation. All participants were offered a standardized breakfast before scanning. Caloric intake was (expectedly) lower in the AN group (**Table 1**). Of the 31 AN, 28 were diagnosed with a restrictive and 3 with a binge-eating/purging subtype.

Paradigm Presentation

The same visual food cues as in the previous study were presented in a block design showing 10 consecutive pictures of food followed by 10 consecutive non-food pictures per block – with a duration of 3 s per picture. As mentioned in Joos et al. (30)

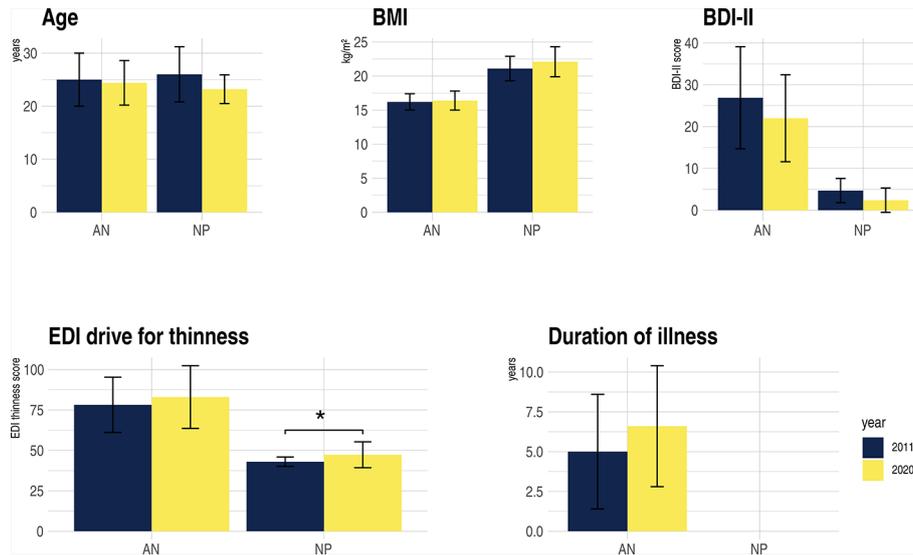


FIGURE 1 | Clinical characteristics of anorexia nervosa (AN) and non-patient (NP), study sample 2011 compared to 2020. BDI-II, Becks Depression-Inventory-2; BMI, Body-Mass-Index; EDI-drive for thinness, Eating Disorder Inventory; kg, kilograms; m², square meter; *p = 0.014. For further clinical characteristics of the replication study see **Table 1**.

TABLE 1 | Clinical characteristics of anorexia nervosa (AN) and non-patients (NP).

	Anorexia Nervosa(N=31)		Non-Patients(N=27)		T-Test	
	Mean	SD	Mean	SD	t-score	p-value
Age (years)	24.0	4.4	23.6	3.0	0.44	0.659
Duration of illness (years)	6.6	3.8	–	–		
Current BMI (kg/m ²)	16.2	1.4	22.1	2.2	-11.97	<.001
Lowest-Lifetime BMI (kg/m ²)	14.8	1.5	20.9	1.8	-11.08	<.001
EDI—total score	61.8	9.3	44.6	3.1	9.19	<.001
EDI—drive for thinness (t values)	83.5	19.6	44.6	6.4	9.85	<.001
EDI—body dissatisfaction (t values)	61.7	12.7	46.6	8.0	5.31	<.001
BDI-II	21.5	10.5	2.3	2.7	9.2	<.001
EDE total score	3.3	1.1	0.4	0.3	13.53	<.001
MWT-B	28.4	5.2	28.0	4.3	0.34	0.736
Caloric intake at breakfast	142.3	157.5	386.3	85.7	-7.12	<.001
STAI-state	38.7	6.6	32.8	4.8	3.83	<.001
STAI-trait	45.5	7.7	29.3	6.8	8.39	<.001

BDI-II, Becks Depression-Inventory-2; BMI, Body-Mass-Index; EDE, Eating Disorder Examination Interview; EDI-2, Eating Disorder Inventory-2; kg, kilogram; m², square meter; MWT-B, Multiple-Choice-Vocabulary-Intelligence Test - German for Mehrfachwahl-Wortschatz-Test-Version (37); SD, standard deviation; STAI, State-Trait Anxiety Inventory.

some of the stimuli have been created by ourselves while others were kindly provided by R. Uher and colleagues (38).

Five blocks of each condition were presented. Examples of the stimuli used can be found in **Supplement 1**.

The instruction was identical to the previous study: participants should watch the pictures attentively (30).

MRI Data Acquisition and Preprocessing

A T1-weighted MPRAGE sequence was recorded as an anatomical reference (repetition time (TR): 2300ms, echo time

(TE): 2.98ms, flip angle (FA): 90°, field-of-view (FOV): 240*256 mm, 176 slices, voxel size: 1x1x1 mm) using a Siemens 3T PRISMA Magnetom (Erlangen, Germany) equipped with a 20-channel head coil. The T1-weighted sequence was followed by the recording of 159 functional echo-planar T2*-weighted (EPI) images (TR: 2,500 ms, TE: 30 ms, FA: 90°, FOV: 192*192 mm, 38 slices, voxel size: 3x3x3 mm, interleaved). All EPI volumes were automatically rigid-body transformed to correct for head motion and a distortion correction algorithm was applied (39).

The statistical parametric mapping software SPM12 [Wellcome Trust Centre of Imaging Neuroscience, London; for details, see (40)] was applied for the preprocessing and statistical analyses of the functional data. The first two volumes of each run were disregarded as so-called dummy scans, an artifact detection algorithm (ArtRepair toolbox, SPM) was applied to detect head motion and spiking artifacts. The realignment to the first volume of the raw functional images that were not motion corrected, was done to generate six head motion parameters (rotation and translation in x, y, z direction). To correct for influences of head motion those parameters were entered in the statistical first-level analysis as regressors of no interest. Using the anatomical MPRAGE image the remaining motion corrected images were spatially normalized with the Montreal National Institute (MNI) reference system followed by the smoothing of the functional images using a three-dimensional isotropic Gaussian kernel (8 mm full width at half maximum) to increase the signal-to-noise ratio and to compensate for inter-individual differences in location of corresponding functional areas. To remove low frequency artifacts across the time-series we applied a high-pass filter (128 s).

Statistical Analyses

Psychometric and behavioral data were assessed by two-sample t-test with a level of significance of $p < 0.05$.

For functional data a linear regression model (general linear model [GLM]) with six regressors, modeling the head motion parameters of the realignment procedure, was fitted to the signal time courses of each voxel for each participant. The food and nonfood regressors were fitted with a canonical hemodynamic response function.

Whole Brain Second Level Analysis Replicating the Original Study

The resulting beta estimates for the two regressors were fed into a voxel-wise group-level random effects analyses using SPM's "full factorial" model with the factors condition (food and nonfood) and group (AN, NP) (30). Two different SPM t-contrasts of differential activation towards food versus nonfood condition were calculated for the comparisons AN(Food>non-food) >/< NP(Food>non-food). Bar graphs of activity were generated using the rfx plot as described by Gläscher (41). For the replication of Joos et al. (30) group activation maps (food versus nonfood) we used for the within-group comparisons a cluster-defining threshold of $p_{\text{uncorr.}} < 0.001$ (> 10 voxels) and for the between-group comparison a cluster-defining threshold of $p_{\text{uncorr.}} < 0.01$ (> 0 voxels). Results were considered significant at $p < 0.05$, corrected for multiple comparisons (Family-wise error corrected (FWE)).

Region of Interest-Based Second Level Analysis Replicating the Original Study

In addition to the whole brain analysis, a region of interest (ROI) approach was conducted. As performed by Joos et al. (30), the following ROIs according to the Automated Anatomical Labeling Atlas [AAL; (42)] were used: medial and lateral orbitofrontal cortex (OFC), amygdala, ACC, insula and parietal

lobe. Again, data were corrected for multiple comparison applying family wise error correction ($p < 0.05$), as a small volume correction (SVC) for all voxels in the corresponding ROI.

Whole Brain Second Level Analysis According to Current Recommendations

Within-group food > nonfood differences were calculated using a one-sample t-test for both the AN and NP group. Further, the food > nonfood contrasts of the two groups were compared in a two-sample t-test. For both analyses the cluster-defining thresholding was set to $p_{\text{uncorr.}} < 0.001$, $k \geq 10$ (43–46).

ROI-Based Second Level Analysis According to Current Recommendations

A SVC was conducted using the ROIs and the t-statistics described above.

Methodological Differences to the Original Study

Sample and State of Participants

The sample size was larger, however clinical characteristics were similar (Figure 1). In the earlier study we neither controlled for menstrual cycle nor hormonal contraception, nor was the breakfast standardized (30). Furthermore, the current study was undertaken in the morning, while the former took place in the afternoon hours.

Paradigm Presentation

Visual stimuli were now presented with a BOLD Screen system, which has a better contrast and resolution than the rear-projection system used in the Joos et al. (30) study. Additionally, other fMRI data were gathered before the food paradigm, which was not the case in the initial study. In the current study, we used the manikins of the International Affective Picture System (47) assessing the emotional response to the visual stimuli after scanning (outside the scanner) in three dimensions (arousal, valence, dominance), as we used this approach with another paradigm (32) as part of the multimodal study. In the previous study the Likert scale was applied.

MRI Data Acquisition and Preprocessing

A comparison of the scanner parameters of the two studies is presented in **Supplement 2**. Due to a scanner upgrade from a Siemens TRIO to a PRISMA system the original MRI parameters could not be adopted. The repetition time (TR) was lowered from 3 to 2.5 s to improve the sampling rate of the BOLD signal. All these changes aimed to increase the signal-to-noise ratio.

Post-processing of the two data sets was always conducted with the SPM standard settings. Yet, there are some differences in the two post-processing pipelines. Joos et al. (30) discarded 10 functional images, while in the current study two dummy scans were discarded in addition to five scans, which were discarded internally by the MR system. In the SPM5 analysis of the initial study the segmentation algorithm for the T1 images differs from the "new segment" procedure used in SPM12, which models the whole head, rather than just the brain. For further details we refer

to “SPM: A history” by J. Ashburner (2012, <https://doi.org/10.1016/j.neuroimage.2011.10.025>).

Statistical Analyses

Additionally to the identical second level and ROI analysis replicating Joos et al. (30) a statistical analysis according to current recommendations was conducted (see *Region of Interest-Based Second Level Analysis Replicating Joos et al. (30)*)

RESULTS

Clinical Characteristics

Clinical details are listed in **Table 1**. The AN and NP group of the current study were of the same age and no significant differences were found in the crystalline intelligence test [MWT-B, (30)]. NP had an expectedly higher BMI than AN. Psychopathology showed typically elevated scores of the questionnaires and interviews in AN (**Table 1**). With respect to the standardized breakfast before the measurement, the AN patients consumed fewer calories than the NP. **Figure 1** illustrates the similarities of

the clinical characteristics of the original compared to the replication study.

Subject Rating of Stimuli

Affective ratings of the food stimuli were more aversive for AN (**Supplement 3**). The AN participants evaluated the food pictures more negatively than the NP in terms of valence, but simultaneously triggered a higher arousal in AN.

Within-Group Activation

In both groups, increased neuronal activity was found in the frontoinsula region and visual cortex observing the food stimuli compared to the neutral stimuli. In addition, AN showed increased activity of the precuneus, supramarginal, postcentral, and angular gyrus and NP of the superior parietal gyrus (**Figure 2A, Supplement 4**).

Group Comparison

Second Level Analysis Replicating the Original Study
Between-group effects yielded higher BOLD signals (AN>NP) in two clusters, one on each hemisphere, including the cingulate

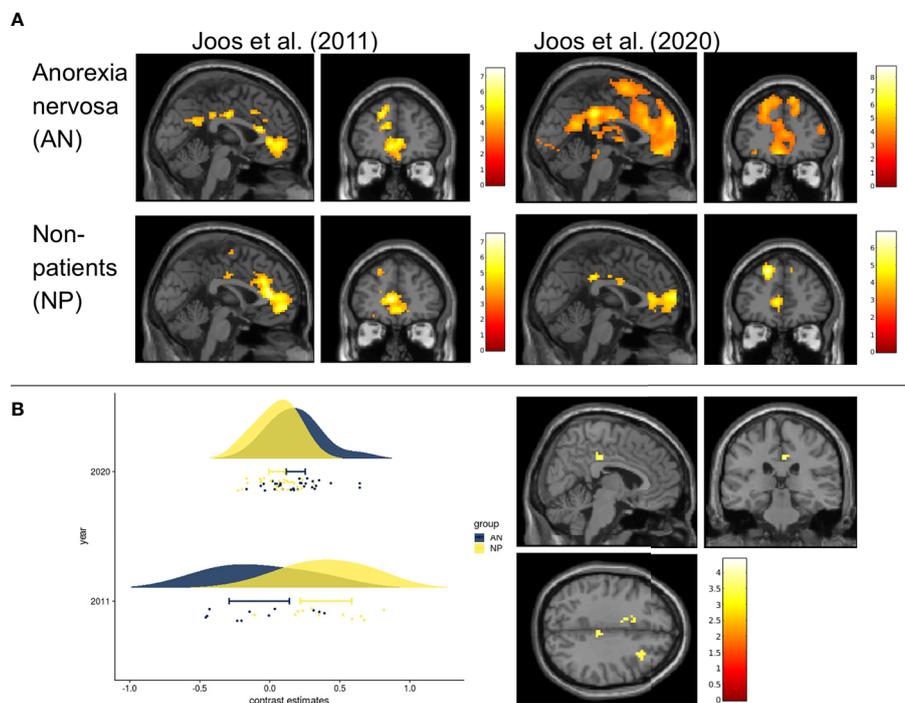


FIGURE 2 | Within group and between group contrasts of the replication study compared to results of Joos et al., 2011 **(A)** Cerebral activation of the within group contrast of anorexia nervosa and non-patients for food>non-food ($p_{\text{uncorrected}} < 0.001$, $k=10$, for visualization purposes). Results of Joos et al., 2020 (one-sample t-test) compared to those of Joos et al., 2011 (full-factorial). All slices at MNI coordinates (0, 45, 0) where chosen as in Joos et al., 2011 for a good comparison. Color bars represent the t-scores (white/yellow = high, red = low). Maps from Joos et al., 2011 with kind permission of Elsevier. **(B)** Left Fig. Rain cloud plot of the contrast estimates (food>non-food) in the non-significant midcingulate cortices (MCC) of the replication study (MNI: $x=6$, $y=-28$, $z=35$, with 3mm radius), next to the MCC which derived from the NP>AN contrast reported by Joos et al., 2011 (MNI: $x=9$, $y=-33$, $z=47$). Right Fig: T-maps of the second level analysis (t-test AN(food>non-food) > NP (food>non-food)) according to the current recommendations (cluster-defining threshold $p < 0.001$). Slices where chosen at non-significant peak cluster activity in the right MCC AN>NP (MNI coordinates $x=6$, $y=-28$, $z=35$); $p_{\text{uncorrected}} < 0.001$, $k > 0$) Color bars represent the t-scores (white/yellow = high, red = low).

cortices, pre-/postcentral gyrus and inferior parietal lobe (IPL) (**Supplement 5**). The contrast NP>AN failed to reveal significant results. In the SVC analyses none of the ROIs showed any group differences.

Second Level Analysis According to Current Recommendations

The two-sample t-test with a threshold of $p_{\text{uncorr.}} < 0.001$ did not yield any between-group effects (**Figure 2B**). Also in the SVC analyses no significant group differences emerged in the ROIs.

DISCUSSION

Our data indicates that within-group effects of food>non-food showed more extensive activation in similar cerebral regions (frontoinsula cortices) in AN and less extensively in NP compared to the previous work (30). Similar patterns of brain activation have been reported in earlier studies that used visual food cues (6). However, when contrasting these activations to NP in the between-group comparison, findings of increased amygdala and decreased MCC activation in AN could not be replicated. In both the current and the previous study (30), as well as in a similar study by Uher et al. (38) AN participants experienced the food stimuli more aversive compared to NP. Therefore, even though the aversive emotions were similar, the neural correlates in the between-group comparison of the studies differed.

The issue of replicability is gaining increased importance in the field of neuroscience, including eating disorders (14, 24, 25). There are several factors that can affect the replicability of results, ranging from the paradigmatic differences to hardware, to intra- and interindividual variances (17). Emotional paradigms seem to be much more critical, particularly in clinical populations (24), which we will discuss in detail below.

In addition to general reasons for poor replicability of studies, such as lack of statistical power, handling of outliers, reporting low p-values or trends (24, 25), and publication biases, the following factors are of particular importance:

1. Compared to within-group statistics, effect-sizes of between-groups in fMRI studies on mental disorders are usually lower (26, 28). From today's point of view, the original study in particular was conducted with a sample size that was too small, which, considering the relatively small effect sizes resulted in a low power of the study. It is therefore likely that the reported results of the original study were false positive or that at least the effect sizes were overestimated, which increases the likelihood of non-replicability. Since the replication study also failed to detect any group differences when applying conservative thresholds, only studies with a large sample size will have enough power to detect the probably rather weak effects. The only way to deal with relatively small effect-sizes is to increase sample size, and efforts such as those of the ENIGMA (Enhancing Neuro Imaging Genetics through Meta-Analysis) consortium pooling data from many sites (17, 25). Furthermore, larger

sample sizes lead to an increase in power (17, 23, 48). As pointed out in several recent papers (43–45, 49), cluster-defining thresholds were often set too low, e.g., $p_{\text{uncorr.}} < 0.01$, which increases the risk of false-positive results. However, this procedure was common at the time of planning the initial study (Woo et al. (44) call it “endemic”). No significant group differences emerged when applying the currently recommended strict thresholds (for further details see, e.g., 42, 43, 44).

2. Heterogeneity across participants is an important confounder, not only in patients but also NP. In our two studies many factors are comparable (age, BMI, duration of disorder, psychopathology, in particular drive for thinness, and most being of the restrictive subtype, depression scores and perception of food pictures are more aversive in AN compared to NP – **Figure 1**), while other confounding genetic, environmental and stochastic factors are difficult or even impossible to account for. Some of these factors likely have larger effect-size than the investigated condition itself (50). Studies with small sample sizes might report results that are based on the effect of uncontrolled variables towards the dependent one (48). This also carries the risk of false-positive results due to sampling error. False-positive results may thus lead into a wrong direction, or even worse, may hinder detecting the real pathophysiological mechanisms (51).
3. Similarly, heterogeneity within participants can impact replicability. Depending on the paradigm, different intrinsic factors can influence the BOLD signal. The current study was controlled for effects of daytime (morning) and state of hunger (standardized meal beforehand), which was not the case in the original study. In the morning, hormonal levels like cortisol are higher; similarly, sex hormones exert cerebral effects (25), which was controlled for in the latter but not in the former study. This also increases the probability of false-positive results of the original study.
4. Heterogeneity across study sites arise from different sources. In addition to different fMRI protocols, scanner hardware and image post-processing pipelines, differences in experimental setup (instructions, interaction with the experimenter, order of tests) have an impact (25). In the current study, participants were subjected to other MRI paradigms before the food paradigm was assessed. In the former study participants started with the food paradigm. While an identical post-processing pipeline was used, fMRI protocols and the scanner hardware differed (see material and methods 2.2., **Supplement 2**). Still, person-related variance seems to be clearly greater than site-related variance (24, 25, 50).

Limitation

The cluster-defining threshold of $p < 0.01$ and the full-factorial model in the between group comparisons are a limitation of the former study. This approach is not in line with the current recommendations. In order to ensure the replication of the former study, we applied a methodology as similar as possible, starting with the same statistical between-group analysis and

followed by a statistical analysis according to the current recommendations. Despite being considerably larger than in the previous study, the sample size was still too small. As recent studies point out, due to low effect sizes in the field of fMRI research sample sizes of 100 (52) or even more participants would be necessary (29) to achieve a sufficient power for many effects. Considering these issues, it will be difficult to recruit enough participants in diseases with low prevalence and often low motivation like AN within single center trials; also, costs and efforts will be very high.

Modern scanner hardware seem to influence variability only modestly (24, 25). Differences between SPM5 and SPM12 are mainly in the improved segmentation process and should explain only a minor part of the variance (53).

Another issue discussed in the literature is temporal and spatial stability of fMRI which is influenced by the sensitivity of detecting short-term metabolic changes and neuromodulatory effects (54). Therefore, Logothetis (54) points towards the fact that the fMRI signal of neuromodulatory effects may exceed the signals of purely task-related neuronal activity. This influences not only temporal but also spatial stability. Furthermore, temporal differences in attention, motivation, and excitement, as well as different cognitive strategies for task accomplishment, or changes in cognitive strategy when working on a task, can significantly influence neural activity in response (24). In the original as well as in the replication study, we performed a cross-sectional analysis with a onetime measurement of the participants. Therefore, we cannot assess the influences of short-term metabolic changes and neuromodulatory effects on the BOLD-signals measured. Especially task fMRI studies and within those particularly clinical populations with emotional paradigms seem to be influenced by temporal and spatial instability (24, 29).

CONCLUSION

In the replication study, we were not able to identify elevated BOLD responses of the right amygdala and decreased activation in midcingulate cortices (MCC) in AN compared to NP in the between-group analysis and therefore could not replicate the original study (30). As expected, we and other authors (24, 25) assume that human influences (inter- and intra-individual variances) are greater than most other factors and more difficult to control, especially in emotional tasks and in clinical populations.

Nevertheless, like most other fMRI studies that examine neural correlation of food compared to non-food stimuli (5–8), we found differences between AN and NP while processing food versus non-food stimuli applying the second level analysis replicating Joos et al. (30). The increased activation in AN>NP in the MCC together with the pre-/postcentral gyrus has also been reported by others: an increased cingulate activation was described by Ellison et al. (4) and Gizewski et al. (55), an pre-/postcentral gyrus activation by Boehm et al. (56). No increased IPL activation has been mentioned in AN, while a decreased IPL

activation could be observed in three studies (38, 57, 58). Of those studies included in the meta-analysis and reviews only Kerr et al. (59) reported no differences between AN and NP for food versus non-food. Due to the heterogeneity of the previous results, no definitive conclusions can yet be drawn from these studies. Further, second level analysis according to current recommendations with a threshold of $p_{\text{uncorr.}} < 0.001$ revealed neither between-group effects in the whole brain nor in the ROI analysis.

We aim to understand the cerebral pathophysiology of AN including the pathological eating behavior and maladaptive eating behavior. For valid and reliable conclusions of functionally altered brain regions, replications of fMRI studies examining neural processing of disease-specific food stimuli are paramount. As noted by others, study protocols as well as samples should be precisely described in order to be able to replicate and disentangle possible influences (17, 21, 24, 25). Likely, replication studies should be performed with larger sample sizes to increase the statistical power (26–28). Additionally, longitudinal studies or studies with repeated sessions of the same participants can be used to create replicability maps (17), which can improve the temporal and spatial stability. Besides the lack of replications, reproductions are necessary as well. Reproduction, i.e., the exact re-analysis of the same data (see *Background*), is a necessary step to establish stable data analysis pipelines and therefore also an important prerequisite for replication studies (60).

The issue of replication has been largely neglected in the past and is now increasingly coming into focus. It is of great importance to carefully control and/or describe modifying factors such as hardware, processing pipelines, statistics, experimental setups and clinical descriptions. Since almost all fMRI studies so far have not undergone replication, the validity of most findings in this field can be challenged.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. T-maps of the within and between group comparisons are available at: <https://identifiers.org/neurovault.image:395600>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics commission of the Albert-Ludwig-University Freiburg (Nr. EK-Freiburg 520/13). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Planning of the study: AJ, LT, and AZ. AJ is principal investigator of the DFG project JO 744-2/1. Recruitment and psychosomatic

assessment: AJ, SM, LH, and AZ. Measurement and data analysis: IH, AJ, SM, LH, KN. Writing: IH, AJ, SM, SS, KN, and DE. Proof reading: AJ, SM, IH, SS, LH, KN, DE, LT, and AZ. All authors contributed to the article and approved the submitted version. They agreed to be accountable for all aspects of the work.

FUNDING

The project was funded by the German Research Foundation (DFG Ref: JO 744-2/1). The article processing charge was funded by the University of Freiburg in the funding program Open Access Publishing.

REFERENCES

- Zipfel S, Giel KE, Bulik CM, Hay P, Schmidt U. Anorexia nervosa: aetiology, assessment, and treatment. *Lancet Psychiatry* (2015) 2:1099–111. doi: 10.1016/S2215-0366(15)00356-9
- Fichter MM, Quadflieg N. Mortality in eating disorders - results of a large prospective clinical longitudinal study. *Int J Eat Disord* (2016) 49:391–401. doi: 10.1002/eat.22501
- Treasure J, Zipfel S, Micali N, Wade T, Stice E, Claudino A, et al. Anorexia nervosa. *Nat Rev Dis Primer* (2015) 1:1–21. doi: 10.1038/nrdp.2015.74
- Ellison Z, Foong J, Howard R, Bullmore E, Williams S, Treasure J. Functional anatomy of calorie fear in anorexia nervosa. *Lancet* (1998) 352:1192. doi: 10.1016/S0140-6736(05)60529-6
- Zhu Y, Hu X, Wang J, Chen J, Guo Q, Li C, et al. Processing of Food, Body and Emotional Stimuli in Anorexia Nervosa: A Systematic Review and Meta-analysis of Functional Magnetic Resonance Imaging Studies. *Eur Eat Disord Rev* (2012) 20:439–50. doi: 10.1002/erv.2197
- García-García I, Narberhaus A, Marqués-Iturria I, Garolera M, Rádoi A, Segura B, et al. Neural Responses to Visual Food Cues: Insights from Functional Magnetic Resonance Imaging. *Eur Eat Disord Rev* (2013) 21:89–98. doi: 10.1002/erv.2216
- Lloyd EC, Steinglass JE. What can food-image tasks teach us about anorexia nervosa? A systematic review. *J Eat Disord* (2018) 6(1):31. doi: 10.1186/s40337-018-0217-z
- Simon JJ, Stopyra MA, Friederich H-C. Neural Processing of Disorder-Related Stimuli in Patients with Anorexia Nervosa: A Narrative Review of Brain Imaging Studies. *J Clin Med* (2019) 8:17. doi: 10.3390/jcm8071047
- Committee on Reproducibility and Replicability in Science, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Nuclear and Radiation Studies Board and Division on Earth and Life Studies, et al. *Reproducibility and Replicability in Science*. Washington, D.C.: National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs (2019). doi: 10.17226/25303
- Gilmore RO, Diaz MT, Wyble BA, Yarkoni. Progress Toward Openness T. Transparency, and Reproducibility in Cognitive Neuroscience. *Ann N. Y. Acad Sci* (2017) 1396:5–18. doi: 10.1111/nyas.13325
- Makel MC, Plucker JA, Hegarty B. Replications in Psychology Research: How Often Do They Really Occur? *Perspect Psychol Sci* (2012) 7:537–42. doi: 10.1177/1745691612460688
- Zwaan RA, Etz A, Lucas RE, Donnellan MB. Making replication mainstream. *Behav Brain Sci* (2018) 41:e120. doi: 10.1017/S0140525X17001972
- Baker M. Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. *Nat News* (2016) 3:452–54. doi: 10.1038/533452a
- Gorgolewski KJ, Poldrack. A Practical Guide for Improving Transparency RA. and Reproducibility in Neuroimaging Research. *PLoS Biol* (2016) 14:e1002506. doi: 10.1371/journal.pbio.1002506
- Kampa M, Sebastian A, Wessa M, Tüscher O, Kalisch R, Yuen K. Replication of fMRI group activations in the neuroimaging battery for the Mainz Resilience Project (MARP). *NeuroImage* (2020) 204:116223. doi: 10.1016/j.neuroimage.2019.116223
- Schmidt S. Shall we Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences. *Rev Gen Psychol* (2009) 13:90–100. doi: 10.1037/a0015108
- Bossier H, Roels SP, Seurinck R, Banaschewski T, Barker GJ, Bokde ALW, et al. The empirical replicability of task-based fMRI as a function of sample size. *NeuroImage* (2020) 212:116601. doi: 10.1016/j.neuroimage.2020.116601
- Barba LA. *Terminologies for Reproducible Research*. Prepr ArXiv180203311 (2018). (Accessed December 2, 2019). Available at: <http://arxiv.org/abs/1802.03311>.
- Patil P, Peng RD, Leek JT. What should we expect when we replicate? A statistical view of replicability in psychological science. *Perspect Psychol Sci J Assoc Psychol Sci* (2016) 11:539–44. doi: 10.1177/1745691616646366
- Klapwijk E, van den Bos W, Tamnes CK, Mills K, Raschle N. Opportunities for increased reproducibility and replicability of developmental cognitive neuroscience. (2019) 1–58. doi: 10.31234/osf.io/xfjzt
- Bollen K, Cacioppo JT, Kaplan RM, Krosnick JA, Olds JL. Social, behavioral, and economic sciences perspectives on robust and reliable science: Report of the Subcommittee on Replicability in Science, Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. *Advis. Comm. Natl Sci Found Dir. Soc Behav Econ. Sci* (2015).
- Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline J-B. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage* (2007) 35:105–20. doi: 10.1016/j.neuroimage.2006.11.054
- Turner BO, Paul EJ, Miller MB, Barbey AK. Small sample sizes reduce the replicability of task-based fMRI studies. *Commun Biol* (2018) 1:1–10. doi: 10.1038/s42003-018-0073-z
- Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? *Ann N. Y. Acad Sci* (2010) 1191:133–55. doi: 10.1111/j.1749-6632.2010.05446.x
- Frank GKW, Favaro A, Marsh R, Ehrlich S, Lawson. Toward valid EA. and reliable brain imaging results in eating disorders. *Int J Eat Disord* (2018) 51:250–61. doi: 10.1002/eat.22829
- Chen G, Taylor PA, Cox RW. *NeuroImage* (2017) 147:952–9. doi: 10.1016/j.neuroimage.2016.09.066
- King JA, Frank GKW, Thompson PM, Ehrlich S. Structural Neuroimaging of Anorexia Nervosa: Future Directions in the Quest for Mechanisms Underlying Dynamic Alterations. *Biol Psychiatry* (2018) 83:224–34. doi: 10.1016/j.biopsych.2017.08.011
- Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci* (2017) 18:115–26. doi: 10.1038/nrn.2016.167
- Elliott ML, Knodt AR, Ireland D, Morris ML, Poulton R, Ramrakha S, et al. Poor test-retest reliability of task-fMRI: New empirical evidence and a meta-analysis. *bioRxiv* (2019) 681700. doi: 10.1101/681700
- Joos AAB, Saum B, van Elst LT, Perlov E, Glauche V, Hartmann A, et al. Amygdala hyperreactivity in restrictive anorexia nervosa. *Psychiatry Res Neuroimaging* (2011) 191:189–95. doi: 10.1016/j.pscychres.2010.11.008

ACKNOWLEDGMENTS

This study was carried out as part of the study DFG (German Research Foundation) of DFG-Grant JO 744-2/1. DE was funded by the Berta-Ottenstein-Programme for Advanced Clinician Scientists, Faculty of Medicine, University of Freiburg.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsy.2020.00777/full#supplementary-material>

31. Maier S, Nickel K, Perlov E, Kukies A, Zeeck A, van Elst LT, et al. Insular Cell Integrity Markers Linked to Weight Concern in Anorexia Nervosa—An MR-Spectroscopy Study. *J Clin Med* (2020) 9:1292. doi: 10.3390/jcm9051292
32. Maier S, Spiegelberg J, van Zutphen L, Zeeck A, van Elst LT, Hartmann A, et al. Neurobiological signature of intimacy in anorexia nervosa. *Eur Eat Disord Rev* (2019) 27:315–22. doi: 10.1002/erv.2663
33. Maier S, Schneider K, Stark C, Zeeck A, Tebartz van Elst L, Holovics L, et al. Fear Network Unresponsiveness in Women with Anorexia Nervosa. *Psychother Psychosom*. (2019) 88:238–40. doi: 10.1159/000495367
34. Nickel K, Joos A, van Elst LT, Holovics L, Endres D, Zeeck A, et al. Altered cortical folding and reduced sulcal depth in adults with anorexia nervosa. *Eur Eat Disord Rev* (2019) 27:655–70. doi: 10.1002/erv.2685
35. Nickel K, Tebartz van Elst L, Holovics L, Feige B, Glauche V, Fortenbacher T, et al. White Matter Abnormalities in the Corpus Callosum in Acute and Recovered Anorexia Nervosa Patients—A Diffusion Tensor Imaging Study. *Front Psychiatry* (2019) 10:490. doi: 10.3389/fpsy.2019.00490
36. Nickel K, Joos A, van Elst LT, Matthis J, Holovics L, Endres D, et al. Recovery of cortical volume and thickness after remission from acute anorexia nervosa. *Int J Eat Disord* (2018) 51:1056–69. doi: 10.1002/eat.22918
37. Lehl S, Triebig G, Fischer B. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurol Scand* (1995) 91:335–45. doi: 10.1111/j.1600-0404.1995.tb07018.x
38. Uher R, Murphy T, Brammer MJ, Dalgleish T, Phillips ML, Ng VW, et al. Medial Prefrontal Cortex Activity Associated With Symptom Provocation in Eating Disorders. *Am J Psychiatry* (2004) 161:1238–46. doi: 10.1176/appi.ajp.161.7.1238
39. Zaitsev M, Hennig J, Speck. Point spread function mapping with parallel imaging techniques O. and high acceleration factors: Fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn. Reson. Med* (2004) 52:1156–66. doi: 10.1002/mrm.20261
40. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. *Hum Brain Mapp*. (1994) 1:153–71. doi: 10.1002/hbm.460010207
41. Gläscher J. Visualization of Group Inference Data in Functional Neuroimaging. *Neuroinformatics* (2009) 7:73–82. doi: 10.1007/s12021-008-9042-x
42. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage* (2002) 15:273–89. doi: 10.1006/nimg.2001.0978
43. Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci* (2016) 113:7900–5. doi: 10.1073/pnas.1602413113
44. Woo C-W, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* (2014) 91:412–9. doi: 10.1016/j.neuroimage.2013.12.058
45. Eklund A, Knutsson H, Nichols TE. Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. *Hum Brain Mapp*. (2019) 40:2017–32. doi: 10.1002/hbm.24350
46. Roiser JP, Linden DE, Gorno-Tempini ML, Moran RJ, Dickerson BC, Grafton ST. Minimum statistical standards for submissions to Neuroimage: Clinical. *NeuroImage Clin* (2016) 12:1045–7. doi: 10.1016/j.nicl.2016.08.002
47. Bradley MM, Lang PJ. International Affective Picture System. In: Zeigler-Hill V, Shackelford TK, editors. *Encyclopedia of Personality and Individual Differences*. Cham: Springer International Publishing. (2017) p. 1–4. doi: 10.1007/978-3-319-28099-8_42-1
48. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* (2013) 14:365–76. doi: 10.1038/nrn3475
49. Cox RW, Chen G, Glen DR, Reynolds RC, Taylor. fMRI clustering PA. and false-positive rates. *Proc Natl Acad Sci* (2017) 114:E3370–1. doi: 10.1073/pnas.1614961114
50. Gee DG, McEwen SC, Forsyth JK, Haut KM, Bearden CE, Addington J, et al. Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study. *Hum Brain Mapp*. (2015) 36:2558–79. doi: 10.1002/hbm.22791
51. Geissberger N, Tik M, Sladky R, Woletz M, Schuler A-L, Willinger D, et al. Reproducibility of amygdala activation in facial emotion processing at 7T. *NeuroImage* (2020) 211:116585. doi: 10.1016/j.neuroimage.2020.116585
52. Geuter S, Qi G, Welsh RC, Wager TD, Lindquist MA. Effect Size and Power in fMRI Group Analysis. *bioRxiv* (2018) 295048. doi: 10.1101/295048
53. Ashburner J, Barnes G, Chen C, Daunizeau J, Flandin G, Friston K, et al. *SPM12 manual*. London, UK: Functional Imaging Laboratory Wellcome Trust Centre for Neuroimaging Institute of Neurology (2014). p. 2464.
54. Logothetis. What we can do NK. and what we cannot do with fMRI. *Nature* (2008) 453:869–78. doi: 10.1038/nature06976
55. Gizewski ER, Rosenberger C, de Greiff A, Moll A, Senf W, Wanke I, et al. Influence of Satiety and Subjective Valence Rating on Cerebral Activation Patterns in Response to Visual Stimulation with High-Calorie Stimuli among Restrictive Anorectic and Control Women. *Neuropsychobiology* (2010) 62:182–92. doi: 10.1159/000319360
56. Boehm I, King JA, Bernardoni F, Geisler D, Seidel M, Ritschel F, et al. Subliminal and supraliminal processing of reward-related stimuli in anorexia nervosa. *Psychol Med* (2018) 48:790–800. doi: 10.1017/S0033291717002161
57. Santel S, Baving L, Krauel K, Münte TF, Rotte. Hunger M. and satiety in anorexia nervosa: fMRI during cognitive processing of food pictures. *Brain Res* (2006) 1114:138–48. doi: 10.1016/j.brainres.2006.07.045
58. Scaife JC, Godier LR, Reinecke A, Harmer CJ, Park RJ. Differential activation of the frontal pole to high vs low calorie foods: The neural basis of food preference in Anorexia Nervosa? *Psychiatry Res* (2016) 258:44–53. doi: 10.1016/j.psychres.2016.10.004
59. Kerr KL, Moseman SE, Avery JA, Bodurka J, Simmons WK. Influence of visceral interoceptive experience on the brain's response to food images in anorexia nervosa. *Psychosom. Med* (2017) 79:777–84. doi: 10.1097/PSY.0000000000000486
60. Asendorpf JB, Conner M, Fruyt FD, Houwer JD, Denissen JJA, Fiedler K, et al. Recommendations for Increasing Replicability in Psychology. *Eur J Pers* (2013) 27:108–19. doi: 10.1002/per.1919

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Horster, Nickel, Holovics, Schmidt, Endres, Tebartz van Elst, Zeeck, Maier and Joos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.