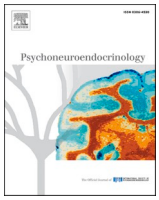
Psychoneuroendocrinology 180 (2025) 107569

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen



A multi-site comparison of the multimodal response to an acute social stressor in the MRI environment

Zala Reppmann [a], Sophie A. Bögemann [b], Netali Mor [c,d], Julian Mituniewicz [e], Natalia Robak [e], Matthias Zerban [f,g], Antje Riepenhausen [a], Lukas Lengersdorff [h], Lara M.C. Puhlmann [g,i], Carolin Wackerhagen [a], Kenneth S.L. Yuen [f,g], Raffael Kalisch [f,g], Dorota Kobylińska [e], Karin Roelofs [j,k], Talma Hendler [c,d,l,m], Erno J. Hermans [b], Henrik Walter [a,n], Ilya M. Veer [a,o,*]

[a] Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[b] Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands
[c] Sagol Brain Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
[d] Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel
[e] Faculty of Psychology, University of Warsaw, Warsaw, Poland
[f] Neuroimaging Center, Focus Program Translational Neuroscience (FTN), Johannes Gutenberg University Medical Center, Mainz, Germany
[g] Leibniz Institute for Resilience Research, Mainz, Germany
[h] Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria
[i] Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany
[j] Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands
[k] Center for Cognitive Neuroimaging, Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands
[l] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
[m] School of Psychological Science, Tel Aviv University, Tel Aviv, Israel
[n] German Center of Mental Health (DZPG), partner site Berlin/Potsdam, Berlin, Germany
[o] Department of Developmental Psychology, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Stress
Multi-site
FMRI
Bayesian
Scan*STRESS*
Resilience

ABSTRACT

Understanding the mechanisms behind (mal)adaptive stress responses is crucial for addressing stress-related mental disorders, which remain leading contributors to global disability and mortality. However, individual differences in stress responses present a challenge for single studies due to limited sample sizes. Multi-site studies can overcome this by increasing statistical power and generalizability, but it remains unclear whether even optimally harmonized procedures can ensure cross-site comparability. To investigate the impact of study site on the multimodal response to an acute social stressor, we analyzed data from the Dynamic Modeling of Resilience Observational (DynaM-OBS) Study, encompassing five study sites across Europe and Israel. By employing harmonized procedures for stress induction through the adapted Scan*STRESS-C* MRI paradigm, along with consistent protocols for data acquisition and processing, we assessed the following markers: subjective stress ratings, heart rate, salivary cortisol, salivary alpha-amylase levels, and fMRI BOLD response. Bayesian inference allowed us to evaluate the evidence for and against the presence of site effects on stress markers. Results indicate successful stress induction, as evidenced by subjective, cardiac, and neural measures, though the salivary stress markers did, on average, not show a typical increase. Comparable stress responses were observed across most sites, highlighting the potential of rigorous procedural harmonization. However, the notable differences at the geographically most distant site may partially reflect variations in stressor exposure, as well as potential cultural differences. These findings highlight the importance of considering demographic and geo-cultural factors in multi-site stress research. Additionally, we emphasize the value of employing Bayesian approaches to integrate and evaluate data from diverse sources. Overall, while such studies enhance statistical power and

* Corresponding author at: Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany.
E-mail address: i.m.veer@uva.nl (I.M. Veer).

https://doi.org/10.1016/j.psyneuen.2025.107569
Received 16 March 2025; Received in revised form 3 August 2025; Accepted 4 August 2025
Available online 5 August 2025

generalizability, careful interpretation of site-specific effects is essential for advancing our understanding of stress-related mental health.

1. Introduction

Stress-related mental disorders, including depression and anxiety, rank among the top 25 leading causes of disability and mortality globally (Vos et al., 2020). This substantial public health challenge highlights the pressing need to deepen our understanding of the mechanisms driving both adaptive and maladaptive stress responses. Acute stress, triggered by perceived threats or environmental disruptions, elicits a complex systemic reaction across multiple physiological systems (Joëls and Baram, 2009). This response can be quantified using various markers, including self-reported assessments, indicators of sympatho-adreno-medullary (SAM) system activation, such as heart rate and salivary alpha-amylase, which typically peaks around 15 min after stress onset (Nater et al., 2006); and measures of hypothalamic-pituitary-adrenal (HPA) axis activity, such as salivary cortisol, which generally peaks 20–30 min after stress exposure (Allen et al., 2014; Man et al., 2023). On the neural level, common activation patterns typically include the insula, premotor cortex, thalamus, fusiform gyrus, amygdala, and inferior prefrontal cortex (Qiu et al., 2022).

However, the stress response can vary significantly from one individual to another and is influenced by factors such as biological sex, menstrual cycle phase, oral contraceptive use (Kajantie and Phillips, 2006; Zänkert et al., 2019), age (Zänkert et al., 2019), aerobic endurance (Wyss et al., 2016), and personality traits such as neuroticism (Kilby et al., 2018), which could hamper the interpretation of findings. Whereas such factors can be carefully controlled for in isolated studies, ample methodological heterogeneity exists between studies, including variations in sample characteristics, study design, and procedures. These variations underscore the challenges in achieving study outcomes that can be compared meaningfully with findings from other studies and complicate efforts to derive generalizable conclusions, especially when aggregating data from different sources (Bayer et al., 2022).

In the stress neuroimaging literature, many different paradigms are used to induce acute stress (Noack et al., 2019), including the Montreal Imaging Stress Test (MIST) (Dedovic et al., 2005), Scan*STRESS* (Streit et al., 2014), Socially Evaluated Cold-Pressor Test (SECPT) (Schwabe et al., 2008; Tutunji et al., 2025; Zhang et al., 2022), and the Aversive Viewing Paradigm (AVP) (van Marle et al., 2009). While these paradigms do typically elicit stress-related SAM and HPA axes responses and the associated neural activation patterns overlap to some extent, differences remain. Some of these differences may result from the distinct stress-inducing components of each task (Berretz et al., 2021; Brunyé et al., 2025). But even when the very same paradigm is used, studies have reported divergent results. For instance, subtle variations in task design, such as differences in the wording of instructions, can substantially affect measures of functional connectivity (Kawagoe et al., 2018).

Multi-site studies are a valuable approach to overcoming the limitations of single-site research by increasing statistical power through the aggregation of data across sites and enabling the detection of subtle effects (Bayer et al., 2022). Additionally, harmonized multi-site study procedures address the issue of variability within a specific paradigm, enhancing the comparability of results within that paradigm. However, even when such harmonized procedures are implemented, it remains unclear whether site-specific factors may still impact results. This issue is particularly relevant in light of the ongoing debate about the need for large sample sizes in reliable (f)MRI analyses (Marek et al., 2022), for which multi-site data collection is indispensable. Resolving this question is critical for determining the reliability of multi-site data aggregation and ensuring that observed differences reflect true biological variability rather than methodological inconsistencies.

In the present study, we address this need by investigating the impact of study site on the response to an acute social stress task in the MRI scanner. To our knowledge, no previous study has systematically examined site effects across multiple sites with identical procedures. We utilized multi-site data from the Dynamic Modeling of Resilience Observational (DynaM-OBS) Study (Wackerhagen et al., 2023). Data were acquired at five different study sites across Europe and Israel. By harmonizing procedures for stress induction, data acquisition, and processing, and by implementing consistent inclusion and exclusion criteria across sites, we have created an optimized framework to assess the impact of study site on markers of the stress response. These markers include subjective stress ratings, levels of salivary cortisol and salivary alpha-amylase, heart rate, and neural activation (fMRI). First, we evaluated the responsiveness of each marker to our stress induction procedure. Next, we investigated the potential effect of study site on each marker. In line with previous research, we hypothesized that the harmonized social stress paradigm would, on average, elicit acute stress responses across participants, as reflected in increased subjective stress ratings, elevated cortisol and alpha-amylase levels, increased heart rate, and characteristic neural activation patterns. We further hypothesized that, due to the rigorous harmonization of procedures across sites, stress responses would not differ systematically between study sites (H0). To this end, we implemented Bayesian inference instead of the predominant frequentist approach. In contrast to the latter, Bayesian statistics allow for quantifying the relative evidence for both H0 and H1. This approach is particularly useful, as it would provide evidence for the validity of aggregating data from multiple sites if H0 holds true (Kelter, 2020). Addressing this gap is crucial for advancing our understanding of the human stress response and improving the robustness and generalizability of findings.

2. Materials and methods

This study used a subset of measures from the DynaM-OBS study related to responses to an acute psychosocial stressor. For full study procedures, see the protocol (Wackerhagen et al., 2023). DynaM-OBS investigates biological, psychological, and social mechanisms of psychological resilience in young adults. Data for the current study were collected between October 2020 and September 2021 at five different sites that collaborated within the EU Horizon 2020-funded DynaMORE project: Charité - Universitätsmedizin Berlin, Department of Psychiatry and Neurosciences in Berlin, Germany; Universitätsmedizin Mainz, Neuroimaging Center (NIC) in Mainz, Germany; Donders Centre for Cognitive Neuroimaging (DCCN) in Nijmegen, The Netherlands; Sagol Brain Institute, Tel Aviv University (TAU) and Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; and University of Warsaw, Faculty of Psychology in Warsaw, Poland. This time period coincided with the COVID-19 pandemic, which affected all sites similarly. Recruitment materials were harmonized across all sites and kept neutral. Procedural harmonization was ensured through standardized instructions, site visits, and ongoing communication (see Supplementary Materials for details).

2.1. Participants

In the DynaM-OBS study, we recruited young adults (aged 18–25) from academic or vocational training settings, leveraging this population's heightened risk for stress-related mental health issues (Gould, 2014; Reavley and Jorm, 2010) to enhance our ability to detect resilience mechanisms. In Israel, the age range was extended to 18–27 to account for study delays due to mandatory military service (32 months for males, 24 for females; (Israel Defense Forces, 2025)). To complement

this age-related vulnerability with additional risk factors, inclusion required ≥ 3 adverse life events (Life Events Inventory, LEI; Cochrane and Robertson, 1973) and elevated psychological distress (General Health Questionnaire score >20; GHQ-28, Goldberg et al., 1997). The Mini-International Neuropsychiatric Interview (M.I.N.I.; Sheehan et al., 1998) was used to confirm the absence of major psychopathology, except for mild depression or tobacco dependence. Of 230 eligible participants with available data on age, gender, and at least one stress marker, seven were excluded due to equipment malfunctions during stress induction, leaving 223 for analysis (see Table 1). Nine more were dropped for incomplete tasks or MRI issues, yielding 214 in the final fMRI sample. All sites received local ethics approval and written informed consent was obtained. Participants were remunerated (details in Wackerhagen et al., 2023).

2.2. Questionnaires

The GHQ-28 and LEI were first administered anonymously during the online pre-screening to assess whether psychological distress and burdensome life events met the study's inclusion criteria, and were reassessed in the week following baseline assessments. Scores here reflect data from month 1, week 1 (i.e., completed within one week following the fMRI session). A trained staff member conducted the M.I.N.I. for on-site screening. Detailed descriptions are in the Supplementary Material. An overview of assessments can be found in Fig. 1A.

2.3. Stress induction

The stress induction paradigm was part of the broader DynaM-OBS neuroimaging battery, following other fMRI tasks (see Fig. 1B). To minimize diurnal fluctuations (Noack et al., 2019), sessions ran between 12:30 and 17:00, with participants instructed to rise at least four hours prior, avoid eating/smoking/caffeine/sugar for two hours, skip exercise that day, and abstain from alcohol for 24 h (Wackerhagen et al., 2023). Before scanning, participants underwent an acclimatization period, including eligibility checks and familiarization with fMRI procedures. Acute social stress was induced using a modification of the Scan*STRESS-C* paradigm (Sandner et al., 2020), which itself is an adapted version of the original Scan*STRESS* task (Streit et al., 2014). Our adaptation retained the core features of Scan*STRESS*-C, including alternating mental rotation and subtraction tasks, a live video feed of the stressor monitoring performance, and other key stress-inducing components such as task difficulty, time constraints (automatically adjusted for poor performance), and negative verbal/visual feedback (see Fig. 1C-D). Unlike Sandner et al. (2020), who presented stress and control blocks in separate runs, we omitted the control run and instead contrasted the stress blocks with the implicit baseline (fixation cross with no cognitive or motor demand). This decision aimed to shorten the paradigm, given the already extensive scanning battery. Unpublished pilot data supported this approach, indicating comparable stress-related neural activation patterns across both contrasts (stress vs. control and stress vs. implicit baseline). Two practice blocks featuring mental rotation and subtraction, along with negative visual feedback, were conducted inside the MRI scanner, though no functional images were acquired. These practice blocks served a dual purpose: to familiarize participants with the task and to provide a reference point for the negative verbal feedback that followed, which criticized their performance and pressured them to improve. Image acquisition began immediately after the verbal feedback was given. All following stress blocks involved only non-verbal stress components during functional imaging. Due to COVID-related restrictions, we deviated from the original setup by employing a single individual to administer the stress task. In most cases, this individual was female (91.5 %), consistent with the predominantly female study staff, wore a white lab coat, and was unfamiliar to the participants.

2.4. Psychological and endocrine stress markers

Subjective stress was assessed between MRI sequences (Fig. 1B) using a 0–10 scale, where 0 indicated no stress and 10 indicated extreme stress. We analyzed ratings taken just before the stress induction (pre-stress) and at three time points afterward (post-stress 1–3), expecting an increase at post-stress 1, assessed ~13 min after stress.

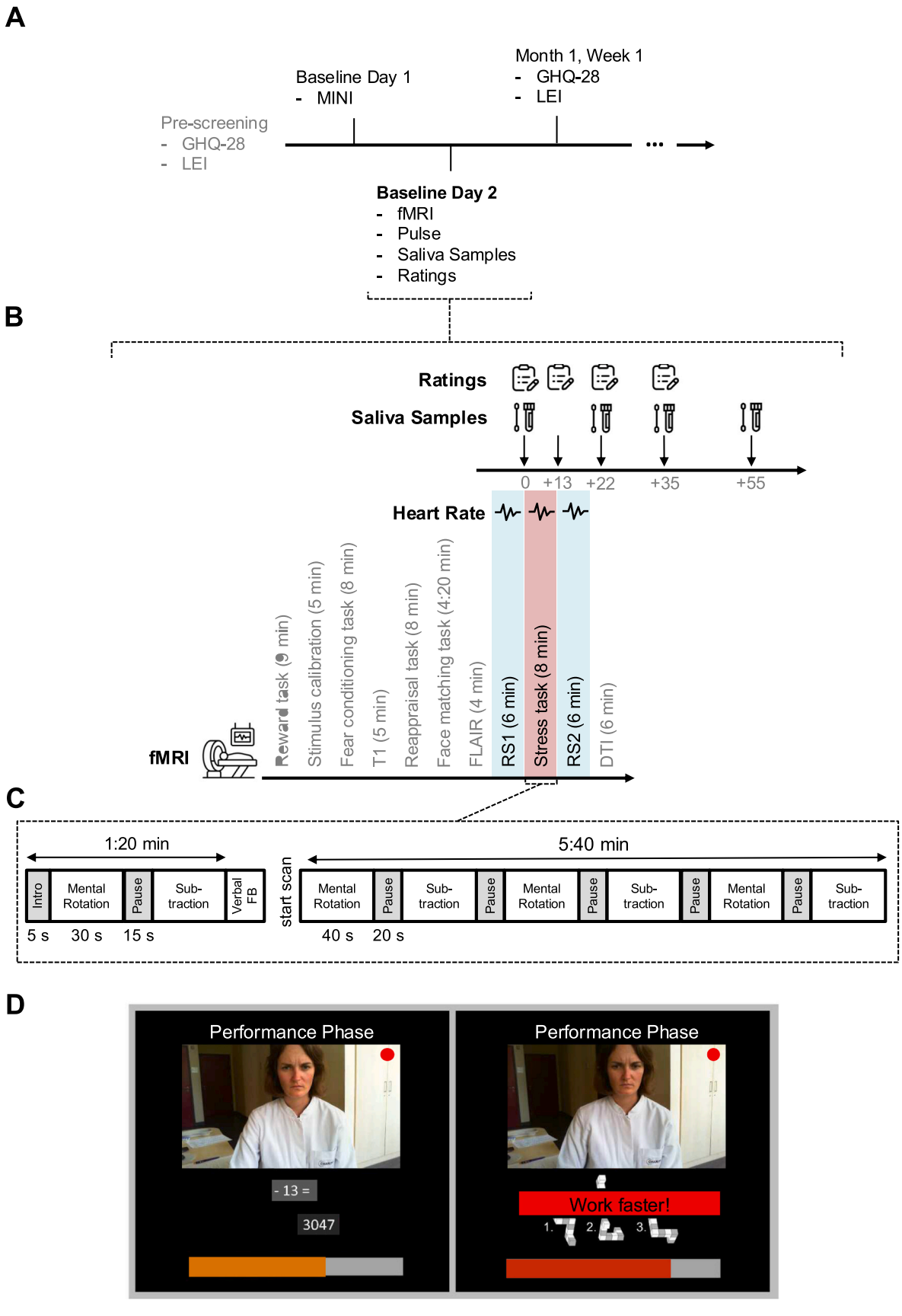
Pulse data were continuously recorded with an MRI-compatible oximeter (Siemens Healthineers, Erlangen, Germany) on the left index finger. We analyzed heart rate from the resting-state scan before stress (RS1), the adapted Scan*STRESS*-C task, and the subsequent resting-state scan (RS2; Fig. 1B). Data were processed in biopeaks (Brammer, 2020), with automated peak detection and manual corrections. The average heart rate for these sequences was calculated using custom Python code. Five heart rate measurements of four different participants–one from RS1, one from RS2, and three from the adapted Scan*STRESS-C* task–were excluded due to physiologically implausible values (i.e., > 200 bpm or < 40 bpm). We expected higher heart rate during stress than pre/post-stress (Man et al., 2023).

A total of nine saliva samples were collected between MRI sequences using Salivettes (Sarstedt, Nümbrecht, Germany). Participants were instructed to place the cotton swab in their mouths without using their fingers. They were then asked to moisten it for one minute before returning it to the plastic tube again without using their hands. For the current analysis, we included only the samples taken immediately before and after the stress induction (one pre-stress and three post-stress; Fig. 1B). The timing of post-stress samples was determined by expected peaks while accommodating the MRI protocol timeline. Based on established cortisol response patterns, we expected cortisol to rise in the first two post-stress samples and alpha-amylase to peak in the first post-stress sample (Nater et al., 2006; Man et al., 2023). Samples from all sites were stored at -20°C and then analyzed in one batch by Dresden Lab Service GmbH for levels of cortisol and alpha-amylase. Cortisol levels were measured using a chemiluminescence immunoassay with high sensitivity (IBL International, Hamburg, Germany). Alpha-amylase levels were measured using an enzyme kinetic method (for details see Rohleder et al., 2006). Both assays had intra- and inter-assay coefficients

Table 1
Sample Characteristics across Study Sites.

	Overall sample	Berlin (DE)	Mainz (DE)	Nijmegen (NL)	Tel Aviv (IL)	Warsaw (PL)
N	223	43	29	55	43	53
Age*	22.09 (2.34)	22.05 (2.17)	21.24 (1.83)	21.04 (2.20)	24.58 (1.88)	21.64 (1.73)
Gender	M: 95 W: 126 (56.5 %) D: 2	M:14 W: 29 (67.4 %) D: 0	M: 12 W: 16 (55.2 %) D: 1	M: 24 W: 31 (56.4 %) D: 0	M: 19 W: 23 (53.5 %) D: 1	M: 26 W: 27 (50.9 %) D: 0
LEI*	8.47 (4.88)	7.61 (4.07)	7.66 (3.85)	6.72 (3.36)	9.58 (5.62)	10.55 (5.77)
GHQ	22.00 (8.02)	20.50 (7.41)	20.76 (8.63)	22.39 (6.52)	21.76 (6.10)	23.68 (10.48)

Note. Summary of sample characteristics including age, sample size (N), mean age, gender distribution (M = Men, W = Women, D = Diverse), and scores from the Life Events Inventory (LEI), and General Health Questionnaire (GHQ). * denotes evidence for differences between at least two study sites as determined by Bayesian ANOVAs. DE = Germany; IL = Israel; NL = The Netherlands; PL = Poland.



(caption on next page)

Fig. 1. Study overview and assessment timeline of stress-related markers in the DynaM-OBS study. (**A**) General study timeline showing relevant parts of the DynaM-OBS study. Participants were screened for eligibility using the Life Events Inventory (LEI) and the General Health Questionnaire (GHQ-28) during pre-screening, and the Mini-International Neuropsychiatric Interview (M.I.N.I.) administered on Baseline Day 1. Participation then started with an extensive baseline characterization, including fMRI and questionnaires. .
(**B**) Schematic of the assessment of stress-related markers during the fMRI session on Baseline Day 2, including four subjective stress ratings, four saliva samples for cortisol and alpha-amylase pre- and post-stress, and the recording of heart rate during resting state scan 1 (RS1), resting state scan 2 (RS2), and the adapted Scan*STRESS*-C paradigm. The average respective times (in min) of assessment of ratings and saliva samples relative to stressor onset are depicted in grey. (**C**) Schematic of the adapted Scan*STRESS*-C paradigm structure. (**D**) Example screenshots from the adapted Scan*STRESS*-C paradigm showing the subtraction task (left) and mental rotation task (right).

of variation below 7 %.

2.5. fMRI data

Except for Warsaw, neuroimaging data were obtained using the same scan sequences on 3 T MAGNETOM Prisma systems (Siemens Healthineers, Erlangen, Germany) equipped with 32-channel head coils (64-channel head coil at our site in Tel Aviv). In Warsaw, a 3 T MAGNETOM Trio system (Siemens, Germany) with adapted scan sequences was utilized. Please refer to the Supplementary Material for a detailed overview of MRI acquisition parameters. fMRI data of all sites were centrally analyzed and underwent harmonized quality checks using HALFpipe (Waller et al., 2022). Preprocessing steps encompassed motion correction, spatial smoothing with a full width at half maximum (FWHM) of 4 mm, denoising based on ICA-AROMA (Pruim et al., 2015), and non-linear normalization to 2 mm MNI standard space (Evans et al., 2012). A Gaussian high-pass filter was applied with a cutoff of 200 s to remove low-frequency drift. The cutoff was determined using a commonly applied rule-of-thumb of 1.5 times the signal period (Smith, 2003). In our experimental design, the signal period of the math and rotation blocks was 120 s each (i.e., 40 s on and 80 s off), yielding a recommended cut-off of 180 s, which was rounded to 200 s to be certain not to remove any task signal. After preprocessing and quality checking of the fMRI data, voxel-wise first-level contrast (beta) images for both stress blocks (mental rotation and mental arithmetic) against the implicit baseline (fixation cross) were calculated using the general linear model in HALFpipe (Waller et al., 2022).

2.6. Statistical analysis

All non-fMRI data were processed in Python (Van Rossum and Drake, 2009). Missing data ranged from 2.69 % to 13.9 % across stress markers and was imputed using Scikit-learn's IterativeImputer (see Supplementary Material for details). Cortisol, alpha-amylase, and heart rate data were log-transformed to correct skewness.

To examine sociodemographic and psychological variations across the five sites, we used contingency tables (for gender) and Bayesian ANOVAs (for age, LEI, and GHQ) in JASP version 0.18.3 (Love et al., 2019), which utilizes the R package BayesFactor (Morey et al., 2016; van den Bergh et al., 2020). While sociodemographic (e.g., age) and procedural factors (e.g., scanner type) may contribute to site differences, we deliberately chose not to include these variables as covariates in the main analyses because they are systematically linked to our study sites. Adjusting for them could therefore obscure genuine site-specific effects that are conceptually meaningful in cross-cultural stress research. The effectiveness of our stress induction procedure and the potential difference between study sites were assessed via separate Bayesian repeated measures ANOVAs for each stress marker (subjective stress, heart rate, cortisol, alpha-amylase) using JASP's default priors (Rouder et al., 2012), which were appropriate given our exploratory research question and absence of strong prior knowledge. To capture the stress-effect, time was used as the within-subject repeated measures factor with four levels (pre-stress, post-stress 1, post-stress 2, post-stress 3) for subjective stress ratings and the salivary markers, and three levels (pre-stress, stress, post-stress) for heart rate (see Fig. 1B). Study site was used as a between-subjects factor with five levels (Berlin, Mainz, Nijmegen, Tel Aviv, Warsaw). To explore potential inter-individual differences in cortisol responses, we applied Miller et al.'s (2013) classification criteria to distinguish cortisol responders from non-responders based on baseline-to-peak increases (Miller et al., 2013). These criteria encompass threshold values for raw cortisol data (absolute increase $\geq$1.54 nmol/l; percentage increase $\geq$15.47 %) and log-transformed values (absolute increase $\geq$0.14 log[nmol/l]). Contingency table analyses were conducted to assess potential site differences in responder/non-responder ratios.

To evaluate the neural response to the adapted Scan*STRESS-C* task across all participants, we performed a second-level whole-brain voxel-wise analysis on first-level beta-weight images using a permutation-based ordinary least squares (OLS) test (2000 permutations). A two-sided one-sample *t*-test was conducted at each voxel to determine whether the mean effect across participants significantly differed from zero. We controlled for multiple comparisons with FDR correction (Benjamini and Hochberg, 1995) at $\alpha = 0.05$, employing a custom pipeline primarily based on nilearn (Abraham et al., 2014) and SciPy (Virtanen et al., 2020).

To assess differences in neural activity between study sites during the adapted Scan*STRESS-C* task, we used the previously calculated first-level beta weights and calculated average beta weights (per participant) across all voxels within each region of the Brainnetome atlas (246 parcellations, Fan et al., 2016). These region of interest (ROI)-based means were then used for site comparisons via Bayesian ANOVAs (Rouder et al., 2012). A second-level contrast image was created to display Bayes factors (BFs) indicating the evidence for and against site differences in each ROI. Here, we chose an ROI-based approach, as voxel-level analyses would yield more granular patterns of Bayesian evidence that would be less intuitive to interpret. This analysis used a custom pipeline primarily utilizing Nilearn (Abraham et al., 2014), SciPy (Virtanen et al., 2020), and the BayesFactor R package (Morey et al., 2016) with default priors (Rouder et al., 2012). Our code is publicly available on GitHub: https://github.com/reppmaz/DynaMOBS_stress_multisite.

Results of the Bayesian analyses are evaluated using the BF, comparing the likelihood of one statistical model against another given the observed data. This approach is pivotal for hypothesis testing, offering a scale to measure the strength of evidence (Morey et al., 2016). Specifically, we report the inclusion BF (BF_{incl}), which evaluates the evidence for incorporating a predictor into a model by contrasting the likelihood of all models incorporating a specific effect with models not including that effect (van den Bergh et al., 2020). Additionally, we report the BF supporting H1, denoted as BF_{10} (Rosenfeld and Olson, 2021). To enhance clarity, BF_{10} is expressed as $Log(BF_{10})$: negative values favor H0, and positive values favor H1. Evidence strength is categorized as none, anecdotal, moderate, strong, very strong, or extreme (Schönbrodt and Wagenmakers, 2018), depending on BF magnitude. A full categorization of BF_{10} and $Log(BF_{10})$ can be found in Table S1 (Supplementary Materials).

3. Results

3.1. Sample characteristics

Descriptive statistics for these and other sample characteristics are

summarized in Table 1. The Contingency table analysis showed no differences in gender ratios between sites $\chi^2(8) = 7.30$, $p = .50$. The Bayesian ANOVA investigating differences in age across study sites demonstrated extreme evidence for age differences (P(M|data) = 1, Log(BF_{10}) = 29.65). Not surprisingly given the study's inclusion criteria, post hoc tests confirmed extreme evidence for age disparities particularly between Tel Aviv and all other sites (vs. Mainz: Log(BF_{10}) = 17.58; vs. Berlin: Log(BF_{10}) = 11.49; vs. Nijmegen: Log(BF_{10}) = 23.47; vs. Warsaw: Log(BF_{10}) = 21.19), with the subsample from Tel Aviv on average being older than all other sites. Additionally, we found anecdotal evidence for an age difference between Berlin and Nijmegen (Log(BF_{10}) = 0.7), with participants in Berlin being slightly older. For the remaining site comparisons, the evidence ranged from anecdotal to moderate for the absence of age differences.

The Bayesian ANOVA on LEI scores provided moderate evidence of variability in the number of burdensome life events reported across sites (P(M|data) = 0.99, Log(BF_{10}) = 4.77). Post hoc tests revealed that the Tel Aviv and Warsaw subsamples reported a higher incidence of burdensome life events compared to other sites. Notably, there was extreme evidence of a disparity between Warsaw and Nijmegen (Log(BF_{10}) = 6.00), very strong evidence between Tel Aviv and Nijmegen (Log(BF_{10}) = 2.66), moderate evidence between Warsaw and Berlin (Log(BF_{10}) = 1.89), and anecdotal evidence between Tel Aviv and Berlin (Log(BF_{10}) = 0.01), and between Warsaw and Mainz (Log(BF_{10}) = 1.04). For the remaining contrasts, evidence ranged from anecdotal to moderate against a site difference in the number of burdensome life events. For the GHQ, the Bayesian ANOVA yielded strong evidence against the presence of site-based differences in scores (P(M|data) = 0.08, Log(BF_{10}) = -2.50), suggesting comparable levels of internalizing symptoms across all five sites. Sample demographics are presented in Table 1.

3.2. Psychological and endocrine stress markers

The complete results of all Bayesian repeated measures ANOVAs are presented in Tables S2 through S7 and Table S14-15 in the Supplementary Material. Sensitivity analyses confirmed the robustness of these findings across alternative prior specifications (see Supplementary Table S16), with evidence patterns remaining consistent across conservative (0.3) to liberal (1.0) prior scales.

3.2.1. Subjective stress ratings

The analysis revealed that the model incorporating time, site, and their interaction was superior for our data (P(M|data) = 0.99, Log(BF_{10}) = 346.54), offering evidence for the influence of time (Log(BF_{incl}) = ∞), site (Log(BF_{incl}) = 4.79), and their interaction (Log(BF_{incl}) = 5.87) on subjective stress ratings. Post hoc comparisons concerning time show extreme evidence for an increase in stress ratings in response to the stress induction, i.e., from pre-stress to post-stress 1 (Log(BF_{10}) = 144.34), suggesting participants (on average) did perceive stress. Site comparisons show anecdotal to very strong evidence for ratings in Tel Aviv differing from all other sites (vs. Berlin: Log(BF_{10}) = 4.05, Mainz: Log(BF_{10}) = 0.71, Nijmegen: Log(BF_{10}) = 3.87) except Warsaw, where there is anecdotal evidence for no difference (Log(BF_{10}) = -0.45), with stress ratings being overall lower in Tel Aviv. Conversely, anecdotal to substantial evidence is seen for the absence of site differences between all other sites. The interaction effect between time and site seems to be driven by Tel Aviv, with ratings pre-stress being similar to all other sites but showing a lesser increase right after stress (see Fig. 2A).

3.2.2. Heart rate

The model incorporating the factors time, site, and their interaction was established as the optimal model for our data (P(M|data) = 1, Log(BF_{10}) = 209.06), demonstrating evidence for the effects of time (Log(BF_{incl}) = 32.75), site (Log(BF_{incl}) = 12.06), and their interaction (Log(BF_{incl}) = 13.70) on heart rate. Post hoc tests reveal extreme evidence for an effect of time, with the elevated heart rate during stress compared to pre- and post-stress (Log(BF_{10}) = 107.12 and Log(BF_{10}) = 101.78, respectively), showing that our stress induction paradigm evoked an autonomic response. Post hoc tests for site show substantial evidence for no difference between study sites. The interaction between time and site seems to be driven by Tel Aviv, where the heart rate is similar to all other sites pre- and post-stress but increases to a lesser extent during stress (see Fig. 2B), in line with the stress ratings.

3.2.3. Cortisol

The model including the factors time and site was identified as the best fit for our data (P(M|data) = 0.95, Log(BF_{10}) = 33.94), indicating evidence for the combined influence of time and site cortisol levels (see Fig. 2C). The analysis of effects provided further evidence for the inclusion of the factors time (Log(BF_{incl}) = 29.86) and site (Log(BF_{incl}) = 3.33). Post hoc tests show moderate to strong evidence for no difference in cortisol levels between the pre-stress sample and the first two post-stress samples (vs. post-stress 1: Log(BF_{10}) = -1.95, vs. post-stress 2: Log(BF_{10}) = -2.11), suggesting the absence of a cortisol response to the stress induction paradigm. Concerning site differences, post hoc comparisons reveal extreme evidence for differences between Tel Aviv and all other sites (vs. Berlin Log(BF_{10}) = 15.51, Mainz: Log(BF_{10}) = 15.02, Nijmegen: Log(BF_{10}) = 25.22, Warsaw: Log(BF_{10}) = 14.98), with overall cortisol levels being lower at the Tel Aviv site. Additionally, post hoc tests show moderate evidence for the absence of a site effect between all other sites except Nijmegen and Warsaw where the evidence for the absence of an effect is only anecdotal (Log(BF_{10}) = -0.83).

Individual responder analysis revealed ~ 32 % of participants responded according to both percentage-based and log-transformed absolute criteria, though only 13 % (N = 29) met the absolute raw threshold - a discrepancy likely reflecting overall low cortisol levels in our sample. Detailed results are provided in Figure S1 (Supplementary Material). The contingency table analyses showed no differences in responder-to-non-responder ratios between sites, regardless of which criterion was used for classification (all $p > .05$). Full results are provided in Tables S8 through S13 (Supplementary Material).

3.2.4. Alpha-amylase

The best model for analyzing the effects on alpha-amylase includes the factors time and site (P(M|data) = 0.96, BF_{10} = 98.37), showing evidence for the effects of time (Log(BF_{incl}) = 33.69), and site (Log(BF_{incl}) = 28.67) on alpha-amylase levels (see Fig. 2D). Post hoc tests for the time factor provide strong evidence for no difference in alpha-amylase levels between the pre-stress sample and the first post-stress sample (Log(BF_{10}) = -2.58). However, extreme evidence for increases in alpha-amylase levels is observed between pre-stress and post-stress 2 (Log(BF_{10}) = 16.98) as well as pre-stress and post-stress 3 (Log(BF_{10}) = 40.55), indicating that alpha-amylase levels were elevated later than expected, specifically around 35–55 min after the stress induction. Post hoc tests also show anecdotal to extreme evidence for differences among various sites, except Berlin and Warsaw (Log(BF_{10}) = -0.95), and Nijmegen and Mainz (Log(BF_{10}) = -1.31).

3.2.5. fMRI

All second-level statistical images are available on Neurovault via this link: http://neurovault.org/collections/YACVEXKC/.

3.2.5.1. Neural response to the adapted ScanSTRESS-C task. Comparing stress blocks against the implicit baseline ($p < .05$, whole-brain FDR corrected) revealed widespread activations and deactivations (see Fig. 3). The largest activation cluster, identified using AtlasReader (Notter et al., 2019), spanned the lateral occipital cortex (including V5/MT), extending into the inferior parietal lobule. A complete list of clusters is provided in Table S17 (Supplementary Material). Overall, the observed BOLD response aligns with the expected pattern, consistent

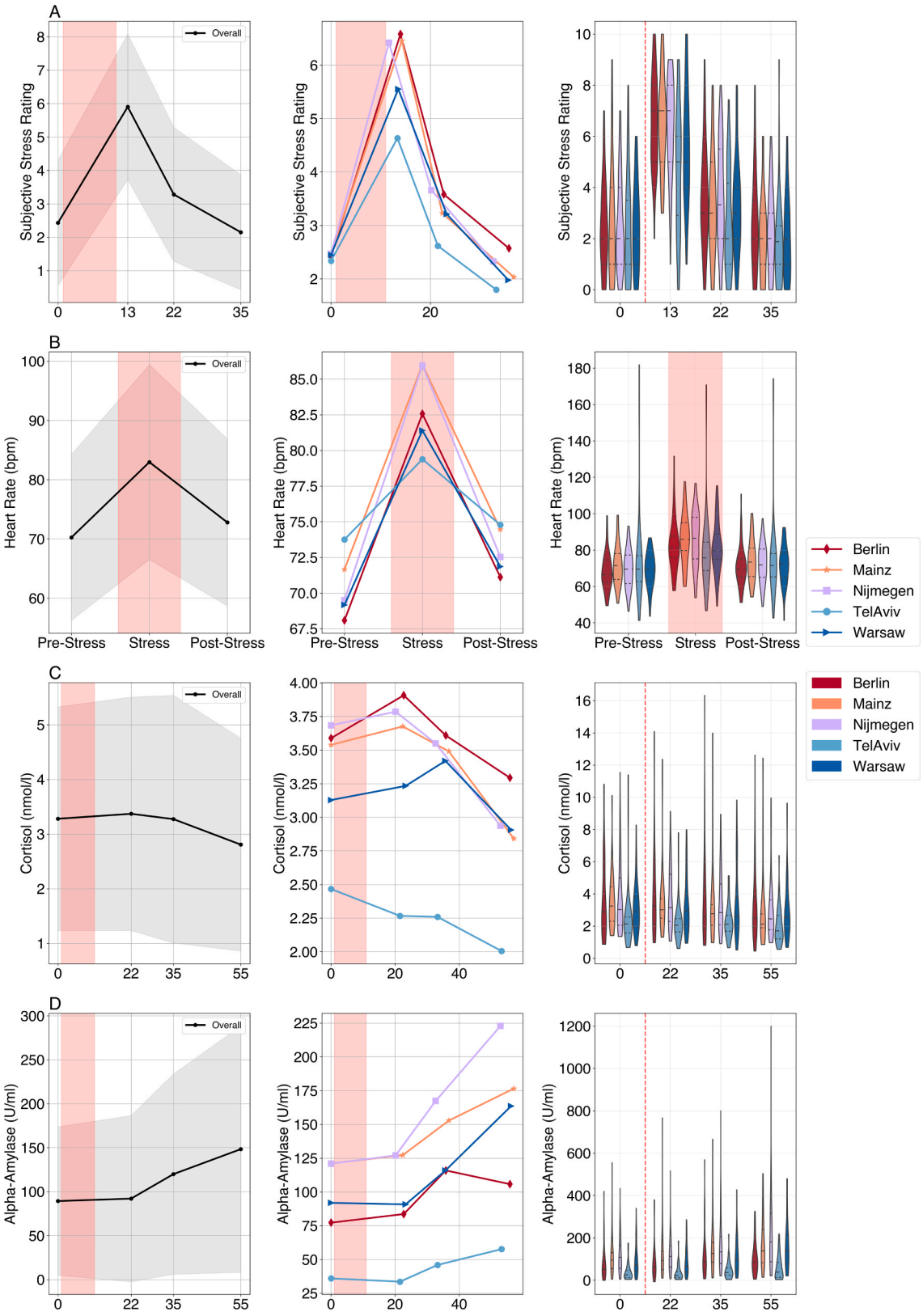


Fig. 2. Multimodal stress response across study sites. Physiological and psychological responses to the adapted Scan*STRESS*-C paradigm shown for the pooled sample (left column) and by study site (middle and right columns). Time courses (left two columns) show minutes relative to stressor onset, with stress induction period highlighted in red. Violin plots (right column) display variability of the means of each site at each sampling point, with stress period marked by vertical red line/shading. Grey shading indicates standard deviation around mean trajectories. **(A)** Subjective stress ratings. **(B)** Heart rate during MRI sequences: pre-stress (resting-state 1), stress (ScanSTRESS-C), and post-stress (resting-state 2). **(C)** Salivary cortisol levels. **(D)** Alpha-amylase levels.

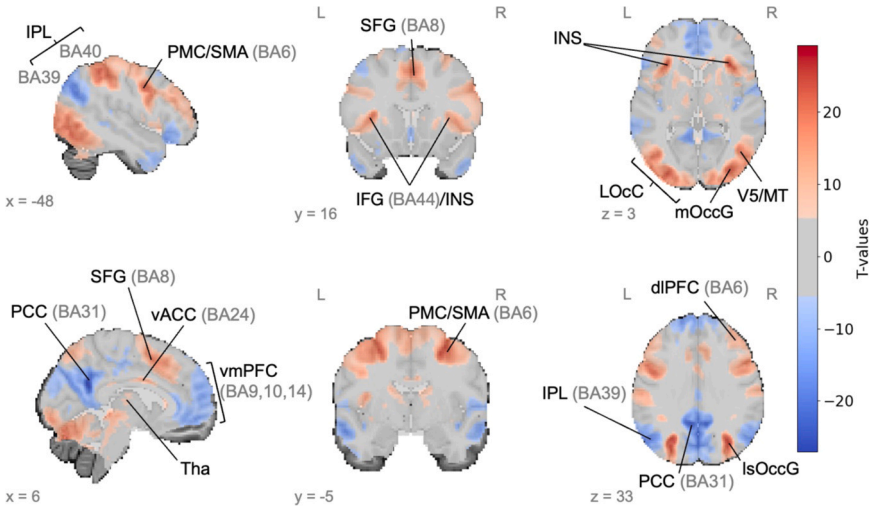


Fig. 3. Neural response to the adapted Scan*STRESS-C* paradigm. Task-specific activations (stress > implicit baseline; red) and deactivations (implicit baseline > stress; blue) are shown ($p < .05$, whole-brain FDR corrected). *T*-values that fall below the minimum absolute *t*-value threshold for significance after FDR correction are shaded in grey, indicating regions where the observed effects were not significant. Region labels are derived from the Brainnetome atlas (Fan et al., 2016). BA = Brodmann area; dla = Dorsal agranular insula; dlPFC = Dorsolateral prefrontal cortex; IFG = Inferior frontal gyrus; INS = Insular cortex; IPL = Inferior parietal lobule; LOoC = Lateral occipital cortex; lsOccG = Lateral superior occipital gyrus; mOccG = Middle occipital gyrus; MVOcC = Medioventral occipital cortex; PCC = Posterior cingulate cortex; PMC = Premotor cortex; SFG = Superior frontal gyrus; SMA = Supplementary motor area; Tha = Thalamus; vACC = ventral anterior cingulate cortex; vmPFC = Ventromedial prefrontal cortex.

with findings from stress studies utilizing the Scan*STRESS-C* (e.g., Sandner et al., 2020) and MIST paradigms (e.g., Dedovic et al., 2005).

3.2.5.2. Differences across study sites in the neural response to the adapted ScanSTRESS-C task. Fig. 4 provides an overview of the differences in neural responses to the adapted Scan*STRESS-C* paradigm between study sites, as identified by the ROI-wise Bayesian ANOVAs. The analysis revealed that ~80 % of ROIs showed evidence for no differences between study sites (H0), as indicated by negative Log(BF_{10}) values, suggesting that participants across sites responded similarly to the stress induction in the vast majority of regions. Evidence for site differences was found in 44 out of all 246 ROIs (see Fig. 4A). Please refer to Table S18 in the Supplementary Material for the full results of these 44 ROIs. To identify task-responsive ROIs, we calculated, within each ROI, the proportion of voxels that exhibited significant activation (stress > implicit baseline or implicit baseline > stress; $p < .05$, whole-brain FDR-corrected). We then ranked all ROIs by the proportion of significantly activated or deactivated voxels and used the 75th percentile (i.e., the top 25 %) as the cutoff, corresponding to a minimum of 66 % significantly active voxels per ROI. Any ROI surpassing this threshold was deemed task-responsive. Out of all 246 ROIs, 62 were deemed task-responsive, of which 16 showed Bayesian evidence for site differences (see red shaded regions in Fig. 4B). The strongest evidence for site differences (H1) among these task-responsive ROIs was found within the ventromedial and lateral occipital cortex (BN193, BN199, BN201, BN205, BN209), the superior frontal gyrus (BN5, BN13), the cingulate gyrus (BN187), and the insular cortex (BN167), the majority within the left hemisphere.

Post hoc tests were performed to identify which specific sites differed from each other among the 44 identified ROIs. The analysis showed that Tel Aviv most frequently differed from the other sites (see Fig. 5A), with ~30 % of the 44 Regions purely driven by differences between Tel Aviv and another site. Overall, the most prominent difference was observed between Tel Aviv and Nijmegen (see Fig. 5B). This indicates that the BOLD responses to the adapted Scan*STRESS-C* paradigm were comparable across all sites, with the exception of Tel Aviv. Full post hoc test results for all 44 ROIs are available in Table S19 (Supplementary Material).

4. Discussion

This study investigated the multimodal response to an acute social stressor and compared it across five study sites from the DynaM-OBS study (Wackerhagen et al., 2023) using Bayesian inference. Stress markers were assessed at various time points before, during, and after the adapted Scan*STRESS-C* task. Samples were demographically similar overall, except that participants in Tel Aviv were slightly older and, together with Warsaw, reported more burdensome life events. The level of internalizing symptoms was comparable across sites.

In examining responses to the adapted Scan*STRESS-C* paradigm, we observed the expected increases in subjective stress ratings and heart rate, indicating that participants indeed experienced stress. While average cortisol levels did not show a pronounced rise following stress induction, approximately one-third of participants (varying by classification criteria) showed a cortisol response. Alpha-amylase exhibited a delayed peak (22 min to 55 min post-stress) instead of the expected peak around 15 min post-stress (Nater et al., 2006). This likely reflects participants getting up from the scanner after an extended period of inactivity, as well as the normal diurnal increases throughout the day (Jantaratnotai et al., 2022), rather than a response to our stess induction. The BOLD response aligned with Sandner et al. (2020) and broader stress literature (Qiu et al., 2022). Overall, our adapted Scan*STRESS-C task* induced stress responses in subjective ratings, heart rate, and neural activity, though the salivary markers did, on average, not show the expected increase.

Among the five sites, Tel Aviv consistently showed lower subjective stress ratings, cortisol levels, and alpha-amylase, along with a smaller heart rate increase, the latter reflecting a difference in stress reactivity. The other four sites appeared largely comparable in their stress responses. Neural analyses likewise revealed strong consistency across sites, with ~80 % of ROIs showing no evidence of site-related differences. However, in the subset of 44 ROIs that did differ, Tel Aviv again emerged as the most distinct, accounting for about a third of those differences. Interestingly, Warsaw, despite using a different MRI scanner, did not show elevated variability, implying that hardware or sequence parameters likely were not the main drivers of site effects. Among the regions identified as task-responsive and showing site-related differences, most were located in the left occipital cortex, raising the

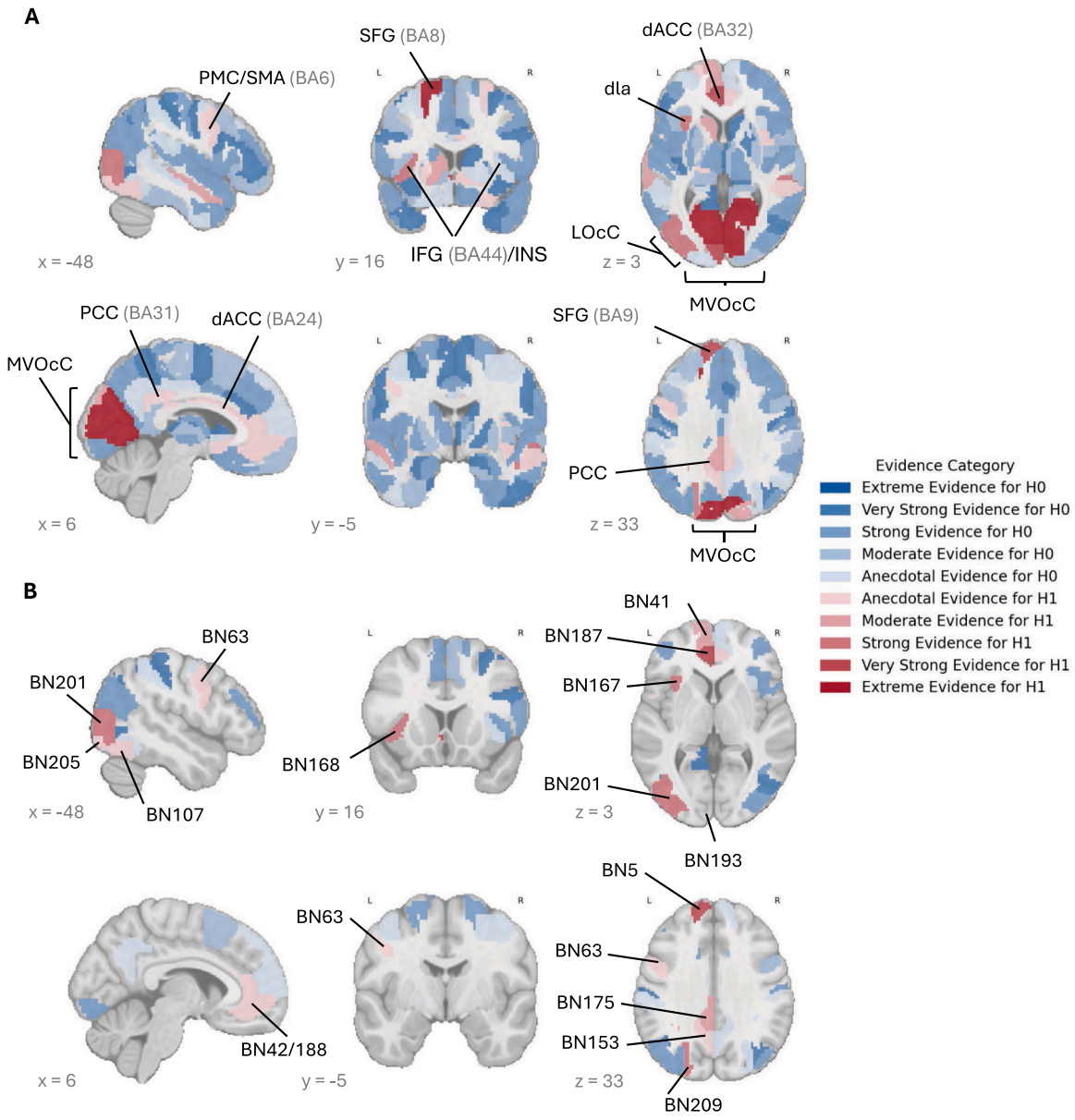


Fig. 4. Color coding is based on Log(BF_{10}), with blue shades representing evidence for no differences between study sites (H0), and red shades indicating evidence for a difference between study sites (H1). Region labels are derived from the Brainnetome atlas (BN, Fan et al., 2016). Labels are shown for regions demonstrating evidence for site differences, with each region labeled only once across panels. BA = Brodmann area; dACC = Dorsal anterior cingulate cortex; dla = Dorsal agranular insula; IFG = Inferior frontal gyrus; INS = Insular cortex; LOoC = Lateral occipital cortex; MVOcC = Medioventral occipital cortex; PCC = Posterior cingulate cortex; PMC = Premotor cortex; SFG = Superior frontal gyrus; SMA = Supplementary motor area.
(A) Bayesian evidence for site differences in the ROI-wise neural response to the adapted Scan*STRESS-C* paradigm within all ROIs. **(B)** Bayesian evidence for site differences in the ROI-wise neural response to the adapted Scan*STRESS-C* paradigm within task-responsive regions only.

possibility that unrecognized variations in lighting conditions potentially influenced visual cortex activity (Mohamed et al., 2002). These observations underscore that, while harmonized protocols can achieve substantial comparability in multi-site stress research, site-specific factors may still shape the stress response.

Various factors may account for the distinct stress responses in Tel Aviv. Although age influences the response to acute stress (Mikneviciute et al., 2023), the 2.46-year age difference on average between Tel Aviv and the other four sites is relatively minor compared to the age differences typically studied in related research and is unlikely to fully account for the site differences in stress responses. A more plausible explanation for the site differences, and in particular the blunted overall levels of cortisol and the blunted cardiovascular stress reactivity observed in the Tel Aviv subsample, may be the cumulative stressor exposure that extends beyond what is captured by the LEI. While participants in Tel Aviv and Warsaw both reported a higher number of burdensome life events, the LEI does not account for site-specific stressors such as political conflict or military service, factors which may be uniquely prevalent and thus underreported, specifically in Tel Aviv. Consequently, the combination of higher LEI scores and unmeasured environmental stressors could explain the altered stress response observed in that subsample. Indeed, repeated or early adversity has been shown to alter HPA axis functioning, sometimes manifesting as a blunted baseline cortisol levels or stress reactivity (Bunea et al., 2017; Wang et al., 2022). Moreover, Xin et al. (2020) found that higher life stress over the past 12 months predicted a blunted heart rate response to acute psychosocial stress, although the authors did not observe an effect on cortisol levels.

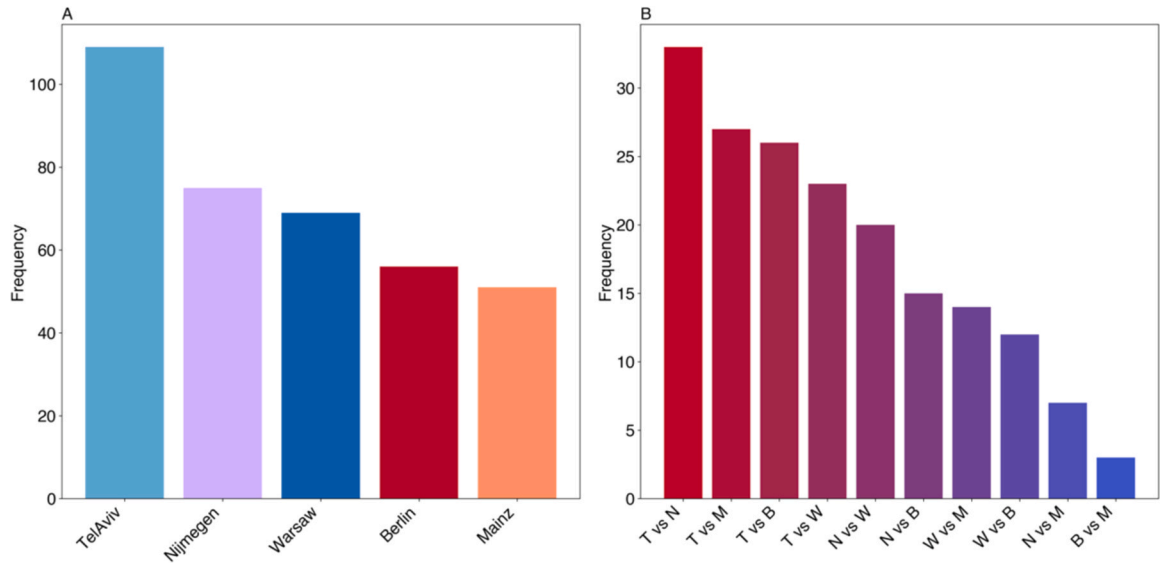


Fig. 5. Frequency of any evidence (anecdotal through extreme) for site differences (H1) in ROI-wise neural response found in the post hoc tests: (**A**) Frequency of ROIs within each study site with evidence for a difference with any of the other sites. (**B**) Frequency of ROIs with evidence for pairwise site differences. B = Berlin; M = Mainz; N = Nijmegen; T = Tel Aviv; W = Warsaw.

Additionally, as the only non-European site, Tel Aviv is geographically the most distant from the four Central/Western European sites and may have more distinct cultural characteristics related to stressor adaptation, potentially contributing to more pronounced differences between Tel Aviv and the European sites compared to the differences among the European sites themselves. The notion that cultural and societal contexts play a crucial role in shaping how individuals react to stress is supported by findings from Miller and Kirschbaum (2019). In their multilevel meta-analysis of 237 studies, approximately 25 % of the variability in cortisol responses to the TSST could be attributed to systematic differences between countries. This finding suggests that unique cultural values, societal contexts, and stressor exposure within a country may partly explain the distinct patterns of stress. While the distinct patterns in Tel Aviv may reflect factors such as mandatory military service or culturally specific stressor appraisal, we did not directly assess these variables, requiring further research.

Although our study highlights Tel Aviv as the most distinct site in terms of stress reactivity, several factors may contribute to site-to-site variation more generally, including differences in age, stressful life events, and cultural contexts. Identifying such differences offers valuable insight into how the interplay of cultural, contextual, and individual-level stressors can lead to heterogeneity in stress responses. Future multi-site investigations could benefit from incorporating site- or culture-specific stress measures. Specifically, we recommend that future studies validate existing stress measures within each cultural context and develop culturally-adapted stress paradigms that incorporate locally-relevant social stressors (e.g., culture-specific evaluation scenarios reflecting local social hierarchies or values), while systematically assessing cultural variables such as collectivism vs. individualism, cultural coping strategies, etc. Additionally, exploring qualitative cultural assessments could provide deeper insights into how cultural meaning-making processes influence stress responses. By doing so, researchers could enhance the comparability of outcomes across diverse settings while providing meaningful knowledge about how environmental and cultural contexts shape stress reactivity.

Several limitations should be acknowledged to contextualize these results and guide future research. Like the MIST (Dedovic et al., 2005) and the Scan*STRESS-C* paradigm (Sandner et al., 2020), our adapted version lacks a tight control condition, making it challenging to distinguish stress-specific from general task-related neural activations. We compared stress blocks to the implicit baseline (i.e., participants looking at a fixation cross without any task or motor demands), which constitutes an even less closely matched control condition than traditionally used. This may have amplified the observed neural differences due to differences in visual stimulation, attentional demands, cognitive load, and motor processing. Additionally, we used only a single person to administer the stress task. This modification, necessitated by COVID-19 safety regulations, may have reduced stress induction intensity. Moreover, the person administering the stress task was female in most cases. While this reflects the composition of the study staff, it represents a potential limitation, as female-only stress panels have been shown to elicit weaker stress responses than male or mixed panels (Goodman et al., 2017). Importantly, this sex composition was not uniform across sites: three sites employed exclusively female stress administrators, while two sites included both male and female staff, with males comprising 13/53 (25 %) in Warsaw and 6/43 (14 %) in Tel Aviv. This unequal distribution across sites could have contributed to between-site differences in stress responses, though notably, the observed pattern of results does not align with the direction predicted by Goodman et al. (2017). The sex of the stress administrator may interact with participants' sexual orientation, an unmeasured variable that could further modulate social-evaluative threat. Furthermore, we did not systematically assess or control for participants' menstrual cycle phase or oral contraceptive use, factors known to influence stress reactivity (Zänkert et al., 2019).

Despite extensive harmonization, we cannot fully dismiss potential site-specific procedural differences. Moreover, our stress induction occurred late in a neuroimaging battery, which may have introduced fatigue, habituation, or emotional carry-over effects (Csathó et al., 2024; Labek and Viviani, 2025; Salihu et al., 2022). While we cannot rule out carry-over effects from preceding tasks, any such influences would be systematic across participants and sites. Additionally, our study sample's demographic characteristics, primarily young, healthy individuals from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) countries, largely recruited from European sites, enhanced comparability across study sites by reducing demographic variability. However, this homogeneity may limit the generalizability of our findings to more diverse populations. This reflects the inherent trade-off between our goal to minimize methodological heterogeneity while assessing site-specific influences on stress responses. Future research aiming to enhance generalizability and ensure sufficient statistical power could benefit from collaborating with large research consortia, such as the

Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium (Thompson et al., 2014), to capture greater demographic variability. Finally, multimodal analytic frameworks, particularly machine-learning approaches that integrate different stress markers, could reveal higher-order patterns not captured by single-modality analyses.

In summary, our findings demonstrate that rigorous procedural harmonization can produce largely comparable stress responses across multiple international sites, underscoring the feasibility of multi-site research on psychosocial stress. However, the distinct pattern observed in Tel Aviv underscores the potential impact of cultural, demographic, and environmental factors, even under controlled conditions. Future investigations should adopt culture-sensitive designs and include contextual measures of adversity and stress to capture the intricate interplay between biology and socio-environmental determinants.

5. Conclusions

This study examined the multimodal response to acute social stress across five international sites to assess comparability and generalizability. Procedural harmonization largely ensured consistency across sites, while the application of Bayesian inference provided a robust framework for investigating site effects on the acute stress response. Nonetheless, notable differences at the only non-European site–possibly linked to slightly higher age, higher stressor exposure, and/or distinct geo-cultural factors–underscore the need to consider demographic, cultural, and societal influences in multi-site studies. Despite inherent challenges, multi-site approaches offer substantial advantages by increasing sample diversity and statistical power, though careful interpretation of site-specific effects is essential for obtaining reliable, generalizable insights into stress-related mental health.

CRediT authorship contribution statement

Zala Reppmann: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Sophie A. Bögemann:** Writing – review & editing, Investigation, Data curation. **Netali Mor:** Writing – review & editing, Investigation, Data curation. **Julian Mituniewicz:** Writing – review & editing, Investigation. **Natalia Robak:** Writing – review & editing, Investigation. **Matthias Zerban:** Writing – review & editing, Investigation, Data curation. **Antje Riepenhausen:** Writing – review & editing, Investigation, Data curation. **Lukas Lengersdorff:** Writing – review & editing, Software, Resources, Methodology. **Lara M. C. Puhlmann:** Writing – review & editing, Investigation, Data curation. **Carolin Wackerhagen:** Writing – review & editing, Investigation. **Kenneth S. L. Yuen:** Writing – review & editing, Conceptualization. **Raffael Kalisch:** Writing – review & editing, Funding acquisition, Conceptualization. **Dorota Kobylińska:** Writing – review & editing, Funding acquisition. **Karin Roelofs:** Writing – review & editing, Funding acquisition. **Talma Hendler:** Writing – review & editing, Funding acquisition, Conceptualization. **Erno J. Hermans:** Writing – review & editing, Funding acquisition, Conceptualization. **Henrik Walter:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Ilya M. Veer:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT, an AI language model, in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

DynaMORE project has received funding from the European Union's Horizon 2020 research under Grant Agreement number 777084. The work of LP was financially supported by the Technische Universität Dresden (Maria Reiche Postdoctoral Fellowship).

We thank S. Stöber and N. Donner for their assistance with project administration. We are also grateful to the following individuals for their essential contributions to the study:

J. Asbreuk, S. Bar, S. Ben-Dor, S. Berman, S. Blum, I. David, D. Even-Or, H. Fiehn, O. Gafni, EJC Hermsen, E. Hodapp, R. Horovich, L. Lagziel, D. Laurila-Epe, L. Knirsch, S. Matys, A. Peschel, J. Piloth, A. C. Rohr, N. Robak, C. Sachs, C. Schultheis, Y. Schwarze, J. Szurnicka, C. Walter, M. Wasylkowska, M. Weber.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2025.107569.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinformatics 8. https://doi.org/10.3389/fninf.2014.00014.

Allen, A.P., Kennedy, P.J., Cryan, J.F., Dinan, T.G., Clarke, G., 2014. Biological and psychological markers of stress in humans: focus on the trier social stress test. Neurosci. Biobehav. Rev. 38, 94–124. https://doi.org/10.1016/j.neubiorev.2013.11.005.

Bayer, J.M.M., Thompson, P.M., Ching, C.R.K., Liu, M., Chen, A., Panzenhagen, A.C., Jahanshad, N., Marquand, A., Schmaal, L., Sämann, P.G., 2022. Site effects how-to and when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. Front. Neurol. 13. ⟨https://www.frontiersin.org/articles/10.3389/fneur.2022.923988⟩.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B (Methodol. 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Berretz, G., Packheiser, J., Kumsta, R., Wolf, O.T., Ocklenburg, S., 2021. The brain under stress–a systematic review and activation likelihood estimation meta-analysis of changes in BOLD signal associated with acute stress exposure. Neurosci. Biobehav. Rev. 124, 89–99. https://doi.org/10.1016/j.neubiorev.2021.01.001.

Brammer, J.C., 2020. Biopeaks: a graphical user interface for feature extraction from heart- and breathing biosignals. J. Open Source Softw. 5 (54), 2621. https://doi.org/10.21105/joss.02621.

Brunyé, T.T., Goring, S.A., Navarro, E., Hart-Pomerantz, H., Grekin, S., McKinlay, A.M., Plessow, F., 2025. Identifying the most effective acute stress induction methods for producing SAM- and HPA-related physiological responses: a meta-analysis. Anxiety Stress Coping 0 (0), 1–23. https://doi.org/10.1080/10615806.2025.2450620.

Bunea, I.M., Szentágotai-Tătar, A., Miu, A.C., 2017. Early-life adversity and cortisol response to social stress: a meta-analysis. Transl. Psychiatry 7 (12), 1–8. https://doi.org/10.1038/s41398-017-0032-3.

Cochrane, R., Robertson, A., 1973. The life events inventory: a measure of the relative severity of psycho-social stressors. J. Psychosom. Res. 17 (2), 135–140. https://doi.org/10.1016/0022-3999(73)90014-7.

Csathó, Á., Van der Linden, D., Matuz, A., 2024. Change in heart rate variability with increasing time-on-task as a marker for mental fatigue: a systematic review. Biol. Psychol. 185, 108727. https://doi.org/10.1016/j.biopsycho.2023.108727.

Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C., 2005. The Montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. J. Psychiatry Neurosci. 30 (5), 319–325.

Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S., 2012. Brain templates and atlases. NeuroImage 62 (2), 911–922. https://doi.org/10.1016/j.neuroimage.2012.01.024.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The human brainnetome Atlas: a new brain Atlas based on connectional architecture. Cereb. Cortex 26 (8), 3508–3526. https://doi.org/10.1093/cercor/bhw157.

Goldberg, D.P., Gater, R., Sartorius, N., Ustun, T.B., Piccinelli, M., Gureje, O., Rutter, C., 1997. The validity of two versions of the GHQ in the WHO study of mental illness in

general health care. Psychol. Med. 27 (1), 191–197. https://doi.org/10.1017/s0033291796004242.
Goodman, W.K., Janson, J., Wolf, J.M., 2017. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the trier social stress test. Psychoneuroendocrinology 80, 26–35. https://doi.org/10.1016/j.psyneuen.2017.02.030.
Gould, J., 2014. Mental health: stressed students reach out for help. Nature 512 (7513), 223–224. https://doi.org/10.1038/nj7513-223a.
Israel Defense Forces. (n.d.). *Our soldiers*. Israel Defense Forces. https://www.idf.il/en/mini-sites/our-soldiers/ (accessed 23 January 2025).
Jantaratnotai, N., Rungnapapaisarn, K., Ratanachamnong, P., Pachimsawat, P., 2022. Comparison of salivary cortisol, amylase, and chromogranin a diurnal profiles in healthy volunteers. Arch. Oral. Biol. 142, 105516. https://doi.org/10.1016/j.archoralbio.2022.105516.
Joëls, M., Baram, T.Z., 2009. The neuro-symphony of stress. Nature Reviews. Neuroscience 10 (6), 459–466. https://doi.org/10.1038/nrn2632.
Kajantie, E., Phillips, D.I.W., 2006. The effects of sex and hormonal status on the physiological response to acute psychosocial stress. Psychoneuroendocrinology 31 (2), 151–178. https://doi.org/10.1016/j.psyneuen.2005.07.002.
Kawagoe, T., Onoda, K., Yamaguchi, S., 2018. Different pre-scanning instructions induce distinct psychological and resting brain states during functional magnetic resonance imaging. Eur. J. Neurosci. 47 (1), 77–82. https://doi.org/10.1111/ejn.13787.
Kelter, R., 2020. Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. BMC Med. Res. Methodol. 20 (1), 142. https://doi.org/10.1186/s12874-020-00980-6.
Kilby, C.J., Sherman, K.A., Wuthrich, V., 2018. Towards understanding interindividual differences in stressor appraisals: a systematic review. Personal. Individ. Differ. 135, 92–100. https://doi.org/10.1016/j.paid.2018.07.001.
Labek, K., Viviani, R., 2025. Functional imaging of time on task and habituation in passive exposure to faces with emotional expression. Neuroreport 36 (3), 135–139. https://doi.org/10.1097/WNR.0000000000002130.
Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q.F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J.N., Morey, R.D., Wagenmakers, E.-J., 2019. JASP: graphical statistical software for common statistical designs. J. Stat. Softw. 88, 1–17. https://doi.org/10.18637/jss.v088.i02.
Man, I.S.C., Shao, R., Hou, W.K., Xin Li, S., Liu, F.Y., Lee, M., Wing, Y.K., Yau, S., Lee, T.M.C., 2023. Multi-systemic evaluation of biological and emotional responses to the trier social stress test: a meta-analysis and systematic review. Front. Neuroendocrinol. 68, 101050. https://doi.org/10.1016/j.yfrne.2022.101050.
van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, van, E.-J., Derks, K., Dablander, F., Gronau, Q.F., Kucharský, Š., Gupta, A.R.K.N., Sarafoglou, A., Voelkel, J.G., Stefan, A., Ly, A., Hinne, M., Matzke, D., Wagenmakers, E.-J., 2020. A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. Annee Psychol. 120 (1), 73–96. https://doi.org/10.3917/anpsy1.201.0073.
van Marle, H.J.F. van, Hermans, E.J., Qin, S., Fernández, G., 2009. From specificity to sensitivity: how acute stress affects amygdala processing of biologically salient stimuli. Biol. Psychiatry 66 (7), 649–655. https://doi.org/10.1016/j.biopsych.2009.05.014.
Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., Malone, S.M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A.M., Earl, E.A., Perrone, A.J., Cordova, M., Doyle, O., Dosenbach, N.U.F., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 603 (7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9.
Mikneviciute, G., Pulopulos, M.M., Allaert, J., Armellini, A., Rimmele, U., Kliegel, M., Ballhausen, N., 2023. Adult age differences in the psychophysiological response to acute stress. Psychoneuroendocrinology 153, 106111. https://doi.org/10.1016/j.psyneuen.2023.106111.
Miller, R., Kirschbaum, C., 2019. Cultures under stress: a cross-national meta-analysis of cortisol responses to the trier social stress test and their association with anxiety-related value orientations and internalizing mental disorders. Psychoneuroendocrinology 105, 147–154. https://doi.org/10.1016/j.psyneuen.2018.12.236.
Miller, R., Plessow, F., Kirschbaum, C., Stalder, T., 2013. Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. Psychosom. Med. 75 (9), 832–840. https://doi.org/10.1097/PSY.0000000000000002.
Mohamed, F.B., Pinus, A.B., Faro, S.H., Patel, D., Tracy, J.I., 2002. BOLD fMRI of the visual cortex: quantitative responses measured with a graded stimulus at 1.5 tesla. J. Magn. Reson. Imaging 16 (2), 128–136. https://doi.org/10.1002/jmri.10155.
Morey, R.D., Romeijn, J.-W., Rouder, J.N., 2016. The philosophy of Bayes factors and the quantification of statistical evidence. J. Math. Psychol. 72, 6–18. https://doi.org/10.1016/j.jmp.2015.11.001.
Nater, U.M., La Marca, R., Florin, L., Moses, A., Langhans, W., Koller, M.M., Ehlert, U., 2006. Stress-induced changes in human salivary alpha-amylase activity–associations with adrenergic activity. Psychoneuroendocrinology 31 (1), 49–58. https://doi.org/10.1016/j.psyneuen.2005.05.010.
Noack, H., Nolte, L., Nieratschker, V., Habel, U., Derntl, B., 2019. Imaging stress: an overview of stress induction methods in the MR scanner. J. Neural Transm. 126 (9), 1187–1202. https://doi.org/10.1007/s00702-018-01965-y.
Notter, M.P., Gale, D., Herholz, P., Markello, R., Notter-Bielser, M.-L., Whitaker, K., 2019. AtlasReader: a python package to generate coordinate tables, region labels, and informative figures from statistical MRI images. J. Open Source Softw. 4 (34), 1257. https://doi.org/10.21105/joss.01257.
Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage 112, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064.
Qiu, Y., Fan, Z., Zhong, M., Yang, J., Wu, K., Huiqing, H., Zhang, R., Guo, Y., Lee, T.M.C., Huang, R., 2022. Brain activation elicited by acute stress: an ALE meta-analysis. Neurosci. Biobehav. Rev. 132, 706–724. https://doi.org/10.1016/j.neubiorev.2021.11.020.
Reavley, N., Jorm, A.F., 2010. Prevention and early intervention to improve mental health in higher education students: a review. Early Interv. Psychiatry 4 (2), 132–142. https://doi.org/10.1111/j.1751-7893.2010.00167.x.
Rohleder, N., Wolf, J.M., Maldonado, E.F., Kirschbaum, C., 2006. The psychosocial stress-induced increase in salivary alpha-amylase is independent of saliva flow rate. Psychophysiology 43 (6), 645–652. https://doi.org/10.1111/j.1469-8986.2006.00457.x.
Rosenfeld, J.P., Olson, J.M., 2021. Bayesian data analysis: a fresh approach to power issues and null hypothesis interpretation. Appl. Psychophysiol. Biofeedback 46 (2), 135–140. https://doi.org/10.1007/s10484-020-09502-y.
Rouder, J.N., Morey, R.D., Speckman, P.L., Province, J.M., 2012. Default Bayes factors for ANOVA designs. J. Math. Psychol. 56 (5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001.
Salihu, A.T., Hill, K.D., Jaberzadeh, S., 2022. Neural mechanisms underlying state mental fatigue: a systematic review and activation likelihood estimation meta-analysis. Rev. Neurosci. 33 (8), 889–917. https://doi.org/10.1515/revneuro-2022-0023.
Sandner, M., Lois, G., Streit, F., Zeier, P., Kirsch, P., Wüst, S., Wessa, M., 2020. Investigating individual stress reactivity: high hair cortisol predicts lower acute stress responses. Psychoneuroendocrinology 118, 104660. https://doi.org/10.1016/j.psyneuen.2020.104660.
Schönbrodt, F.D., Wagenmakers, E.-J., 2018. Bayes factor design analysis: Planning for compelling evidence. Psychon. Bull. Rev. 25 (1), 128–142. https://doi.org/10.3758/s13423-017-1230-y.
Schwabe, L., Haddad, L., Schachinger, H., 2008. HPA axis activation by a socially evaluated cold-pressor test. Psychoneuroendocrinology 33 (6), 890–895. https://doi.org/10.1016/j.psyneuen.2008.03.001.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (20), 11980.
Smith, S.M., 2003. Preparing fMRI data for statistical analysis. In: Jezzard, In.P., Matthews, P.M., Smith, S.M. (Eds.), Functional MRI: An introduction to methods. Oxford University Press, pp. 229–241.
Streit, F., Haddad, L., Paul, T., Frank, J., Schäfer, A., Nikitopoulos, J., Akdeniz, C., Lederbogen, F., Treutlein, J., Witt, S., Meyer-Lindenberg, A., Rietschel, M., Kirsch, P., Wüst, S., 2014. A functional variant in the neuropeptide s receptor 1 gene moderates the influence of urban upbringing on stress processing in the amygdala. Stress 17 (4), 352–361. https://doi.org/10.3109/10253890.2014.921903.
Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M.J., Martin, N.G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N.C., the Alzheimer's Disease Neuroimaging Initiative, E.C., Consortium, I.M.A.G.E.N., Saguenay Youth Study, (S.Y.S.), Group, 2014. The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. 8 (2), 153–182. https://doi.org/10.1007/s11682-013-9269-5.
Tutunji, R., Krentz, M., Kogias, N., Voogd, L. de, Krause, F., Vassena, E., Hermans, E.J., 2025. Changes in large-scale neural networks under stress are linked to affective reactivity to stress in real life. eLife 14. https://doi.org/10.7554/eLife.102574.1.
Van Rossum, G., Drake, F.L., 2009. Python 3 reference manual. CreateSpace.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17 (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.
Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhassani, H., Aboyans, V., Abrams, E.M., Abreu, L.G., Abrigo, M.R.M., Abu-Raddad, L.J., Abushouk, A.I., Murray, C.J.L., 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396 (10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9.
Wackerhagen, C., Veer, I.M., Leeuwen, J.M.C. van, Reppmann, Z., Riepenhausen, A., Bögemann, S.A., Mor, N., Puhlmann, L.M.C., Uściłko, A., Zerban, M., Mituniewicz, J., Lerner, A., Yuen, K.S.L., Köber, G., Marciniak, M.A., Pooseh, S., Weermeijer, J., Arias-Vásquez, A., Binder, H., Walter, H., 2023. Dynamic modelling of mental resilience in young adults: protocol for a longitudinal observational study (DynaM-OBS). JMIR Res. Protoc. 12 (1), e39817. https://doi.org/10.2196/39817.
Waller, L., Erk, S., Pozzi, E., Toenders, Y.J., Haswell, C.C., Büttner, M., Thompson, P.M., Schmaal, L., Morey, R.A., Walter, H., Veer, I.M., 2022. ENIGMA HALFpipe: interactive, reproducible, and efficient analysis for resting-state and task-based fMRI data. Hum. Brain Mapp. 43 (9), 2727–2742. https://doi.org/10.1002/hbm.25829.
Wang, H., van Leeuwen, J.M.C., de Voogd, L.D., Verkes, R.-J., Roozendaal, B., Fernández, G., Hermans, E.J., 2022. Mild early-life stress exaggerates the impact of acute stress on corticolimbic resting-state functional connectivity. Eur. J. Neurosci. 55 (9–10), 2122–2141. https://doi.org/10.1111/ejn.15538.

Wyss, T., Boesch, M., Roos, L., Tschopp, C., Frei, K.M., Annen, H., La Marca, R., 2016. Aerobic fitness level affects cardiovascular and salivary alpha amylase responses to acute psychosocial stress. Sports Med. Open 2 (1), 33. https://doi.org/10.1186/s40798-016-0057-9.

Xin, Y., Yao, Z., Wang, W., Luo, Y., Aleman, A., Wu, J., 2020. Recent life stress predicts blunted acute stress response and the role of executive control. Stress 23 (3), 359–367. https://doi.org/10.1080/10253890.2019.1687684.

Zänkert, S., Bellingrath, S., Wüst, S., Kudielka, B.M., 2019. HPA axis responses to psychological challenge linking stress and disease: what do we know on sources of intra- and interindividual variability? Psychoneuroendocrinology 105, 86–97. https://doi.org/10.1016/j.psyneuen.2018.10.027.

Zhang, W., Kaldewaij, R., Hashemi, M.M., Koch, S.B.J., Smit, A., van Ast, V.A., Beckmann, C.F., Klumpers, F., Roelofs, K., 2022. Acute-stress-induced change in salience network coupling prospectively predicts post-trauma symptom development. Transl. Psychiatry 12 (1), 63. https://doi.org/10.1038/s41398-022-01798-0.