SUPPLEMENTARY MATERIAL

1. SUPPLEMENTARY METHODS

1.1. Measures

1.1.1. Questionnaires

1.1.1.1. Not in booklets

We will use the 17-item *Hamilton Depression Rating Scale (HDRS)* and the *Dysfunctional Attitude Scale (DAS)* at various moments during our study procedure (see main article).

- The *HDRS* is an observer rated major depressive disorder (MDD)-symptom scale to assess the severity of depression.[1] Total scores of the 17-item version range from 0-52 and scores of 0-7 are considered within the normal range; scores of 8-13 indicate mild depression; 13-18 moderate depression; 18-22 severe depression and scores ≥23 indicate very severe MDD. Its internal consistency is high, with Cronbach's α=.80.[2]
- The *DAS* is a self-rated questionnaire to assess general, deeply held, dysfunctional beliefs. Two 40-item versions exists for the Dutch language. Total scores range from 40 to 280, with higher scores reflecting greater dysfunctional attitudes. The internal consistency and test-retest reliability are high, with Cronbach's α's of respectively .90 and .73.[3]

1.1.1.2. Questionnaire-booklet I

We will mail questionnaire-booklet I after the initial assessment, participants will take it with them at the first study-session baseline visit (same for follow-up repeated measure), see main manuscript. It includes the following questionnaires:

- *Inventory of Depressive Symptomatology self-rated (IDS-SR)*: self-rated MDD-symptom scale to assess severity of depressive symptoms. The IDS-SR comprises three factors: cognition and mood, anxiety and arousal, and sleep and appetite regulation. The IDS-SR has 30-items with a total score range from 0 to 84, with higher scores indicating greater severity of depression. Scores ≤13 are considered within the normal range; scores of 14-21 indicate mild MDD; 22-38 moderate MDD; ≥39 severe MDD. The IDS-SR has highly acceptable psychometric properties. Internal consistency was up to 0.92 (Cronbach's α).[4]
- *Leiden Index of Depression Sensitivity-Revised (LEIDS-R)*: self-report questionnaire that measures cognitive reactivity to sad mood.[5] Participants are instructed to think about the last time they felt "somewhat sad", and to indicate - on a 5-point Likert scale ranging from 'not at all' (0) to 'very strongly' (4) - the degree to which a list of statements describe their typical cognitions and

behaviours in response to sad mood. The LEIDS-R contains 34 items which sum up to the total score, and subscales (assessing cognitive reactivity in relation to aggression, hopelessness/suicidality, acceptance/coping, control/perfectionism, risk aversion, and rumination on sadness). The LEIDS-R has good internal consistency (Cronbach's α=0.88).[6]

- *Neuroticism-Extraversion-Openness Five-Factor Inventory (NEO-FFI)*: self-rated questionnaire measuring five factor model ('big-five') dimensions of personality characteristics: Neuroticism, Agreeableness, Conscientiousness, Extraversion and Openness.[7] The NEO-FFI has 60 items on a five point scale, ranging from "strongly disagree" to "strongly agree". It has sufficient internal reliability and two-week retest reliability is uniformly high, ranging from 0.86 to 0.90 for the five scales.[8]
- *Everyday Problem Checklist*:[9 10] Dutch translation of self-rated questionnaire measuring everyday (small) stressors (Daily hassles). The questionnaire consists of 114-items which describe problematic situations and events in daily life.
- *Utrecht Coping List (UCL)*: self-rated measure of coping behaviour while confronted with problems. The questionnaire consists of 47-item on seven empirically derived subscales: active tackling, seeking social support, palliative reacting, avoiding, passive reacting, reassuring thoughts and expression of emotions. The UCL demonstrated strong internal consistency in a study within the UK population. Five of the seven subscales had good test-retest reliability.[11]
- *Negative life events questionnaire*:[10 12] self-rated questionnaire asking for recent life-events of the subject or significant others.
- *Childhood Trauma Questionnaire*:[13 14] Dutch translation of the Childhood Trauma Questionnaire for assessment of childhood adversity. This 28-item self-report questionnaire retrospectively assesses childhood trauma and neglect, and consist of five factors; emotional abuse, emotional neglect, sexual abuse, physical abuse, and physical neglect. Inter-correlations among the five factors ranged from *r*=.41 to *r*=.95.[15] Psychometric properties in a sample of Dutch female sex workers were good.[13]
- *International Physical Activity Questionnair*e (Long Form):[16-18] Dutch translation of an internationally validated questionnaire to measure physical activity, developed with support from the World Health Organization. This 27-item self-report questionnaire assesses the time that participants spent being physically active in the last 7 days. The reliability of the Dutch version was good (intra-class correlation coefficient=0.70-0.96) and the validity moderate (*r*=0.36-0.49) compared to an accelerometer. Reliability and validity is comparable in psychiatric populations, e.g. schizophrenia.[19]

- *Helius Food Frequency Questionnaire (Helius-FFQ)*:[20 21] validated and updated questionnaire to assess dietary intake. The Helius-FFQ enables detailed, standardized, and comparable assessment of the diet of five different ethnic groups.
- *Women's Health Initiative Insomnia Rating Scale (WHIIRS)*:[22] reliable and internationally validated questionnaire to assess sleep quality. WHIIRS' internal consistency is good (Cronbach's α=.78.). Test-retest reliability coefficients were .96 for same-day administration and .66 after a year or more. The WHIIRS has good construct validity.[22]
- *Edinburgh handedness questionnaire*:[23 24] self-rated questionnaire developed to determine dominant laterality in executive functions. The questionnaire assesses handedness by of the preferred hand for carrying out common activities. The 4-item revised structure showed very high reliability on measures of factorial composite reliability and Cronbach's α. Furthermore, the estimate of the quality of factor scores was high.[24]

1.1.1.3. Questionnaire-booklet II

We will hand out questionnaire-booklet II during the first study-session, participants will complete it before the MRI-session (same for follow-up repeated measure). It includes the following questionnaires:

- *IDS-SR* (see above)
- *Snaith-Hamilton Pleasure Scale (SHAPS)*:[25] self-rated questionnaire measuring anhedonia. The SHAPS has 14 items with scores ranging from 0-14. The internal consistency and test-retest reliability of the SHAPS were adequate, with Cronbach's α's of .91 and 0.70 respectively. Furthermore, the SHAPS was significantly correlated with other validated measures of affect and personality.
- *Ruminative Response Scale (RRS-NL)*:[26 27] validated Dutch adaptation of a self-report rumination measure. It consists of 26 items that describe responses to a depressed mood that are focused on the self, symptoms, or consequences of depressed mood. Two separate subscales reflecting pondering and brooding are distinguished. The RRS-NL possesses good internal consistency and validity.[26] In a recent study examining the Dutch version, Cronbach's α for the total RRS-NL was .94, and .64 for the brooding subscale.[28] We adjusted the RRS-NL by slightly reframing the introductory statement. Instead of referring to what subjects generally do when they feel depressed, we asked for their answers reflecting the last week. With this adjustment, we aimed to increase the temporal specificity, by specifically asking for current rumination instead of general ruminative traits.
- *Spielberger State and trait Anxiety Inventory form Dutch Y (STAI-DY)*:[29] self-rated questionnaire measuring state and trait anxiety. The State Anxiety Scale (40 items) measures subjective feelings

of apprehension, tension, nervousness, worry, and activation arousal of the autonomic nervous system. The Trait Anxiety Scale (20 items) evaluates stable aspects of anxiety proneness. Test–retest reliability coefficients ranged from 0.31 to 0.86 and internal consistency was high, ranging from 0.86 to 0.95 (Cronbach's α).[29]

- *Mood and Anxiety Symptoms Questionnaire (MASQ-30D)*:[30] validated short adaptation of the MASQ, designed to measure the dimensions of Clark and Watson's tripartite model in large-scale psychopathology research. The MASQ-30D contains 30 items examining mood and anxiety and has 3 subscales. The scales of the MASQ-D30 showed good internal consistency, with Cronbach's α's >0.87 in patient samples. Correlations of subscales with other measures of mood and anxiety indicated sufficient convergent validity.

1.1.1.4. Questionnaire-booklet III

We will complete questionnaire-booklet III at the MRI-session (same for follow-up repeated measure). It includes three observer-rated questionnaires:

- *HDRS* (see above)
- *CORE checklist for psychomotor MDD-symptoms (CORE)*:[31-33] distinguishes dimensions of psychomotor dysfunction in MDD, all suggestive of a melancholic MDD-subtype. The CORE index is composed of 18 items, scored on a 4-point scale. Factor analysis showed three interpretable domains: (I) retardation items (52% of variance), (II) agitation items (15% of variance), and (III) non-inter-activeness (5% of variance).[34] Its translation in Dutch was recently validated.[35]
- *Salpêtrière retardation rating scale (SRRS)*:[36] measures cognitive and motor aspects of retardation. This scale contains 15 items and has a three-factor solution measuring movement, speech and cognitive function.[34] Correlations between SRRS and measures of cognitive function and motor abilities show good convergent validity.[37]

1.1.1.5. Questionnaire-booklet IV

We will use questionnaire-booklet IV during the follow-up measurements. It consists of the IDS-SR, RRS-NL, and SHAPS, all described above.

1.1.2. Neuropsychological tests

1.1.2.1. Exogenous cueing task (15min)[38]

In this reaction time task, a target stimulus appears at one of two spatial locations, cued by an emotional stimulus (emotional face) preceding at the same ('valid trial') or opposite spatial location of the target ('invalid trial'). When the interval between cue and target onset is short (stimulus onset asynchrony<300ms), participants typically respond faster to valid compared to invalid trials ('cue

validity effect'). In case of emotionally relevant stimulus material, the time course of the cue validity effect may be extended ('enhanced cue validity effect'), leading to a larger cue validity effect compared with neutral information. Secondly, we will measure emotional modification of attentional engagement and disengagement by comparing speed of responding on valid and invalid emotional versus neutral trials. Cue emotional valence may (I) lead to response benefits on valid emotional versus valid neutral trials, which is a measure of attentional engagement towards emotional cues, and/or (II) delay disengagement of attention, which is indexed by a slower reaction on invalid emotional trials compared to neutral trials.[39-41]

1.1.2.2. Facial expression recognition task (20min)

We will use six morphed basic emotions (happiness, surprise, sadness, fear, anger, disgust) from 10 individual characters from the pictures of facial affect series between each prototype and neutral, and present them in random order for 500ms, replaced by a blank screen. We will record reaction times to these emotions and the recognition threshold (the intensity level required for successful recognition of each emotion). This recognition threshold is defined as the level of emotional intensity at which participants correctly identify ≥75% facial expressions of emotion for four consecutive intensities.[42]

1.1.2.3. Emotional categorization (6min)

We will present sixty personality characteristics selected to be disagreeable or agreeable on the computer screen for 500msec each. These words have been translated from the original English version to Dutch, matched in terms of word length, ratings of usage frequency, and meaningfulness. We will ask subjects to categorize the words as likable or dislikeable as quickly and as accurately as possible. Specifically, we will ask to imagine whether they would be pleased or upset if they overheard someone else referring to them as possessing this characteristic, so that the judgment is in part self-referring. We will calculate classification-rates and reaction times to likable and dislikeable words.[40 42]

1.1.2.4. Emotional memory (~5 minutes)

Fifteen minutes after completion of the emotional categorization task, we will ask participants to recall as many of the personality traits as possible. We will compute numbers of positive and negative words recalled for both correct and false responses.[40 42]

1.1.2.5. Internal shift task (12min)[43]

Examines capacity to shift attention between contents of working memory in response to emotional as well as non-emotional material. We will present Karolinska faces at the centre of the computer

screen one at a time. We will ask participants to perform two conditions, a *non-emotional* and an *emotional* one. In the non-emotional condition, we will instruct participants to focus on the relevant stimulus dimension 'gender' (male or female), in the second condition, they have to focus on the 'emotion' dimension (neutral or angry). All participants complete both conditions in counterbalanced order. Participant's task is to keep a silent mental count of the number of items in each category, presented over a block of items (with random 10 to 14 items) and report numbers at the end of each block. We will ask participants to update counters of both categories when a face is presented and report numbers of items at the end of a block in a fixed order, to encourage a consistent counting strategy (e.g. neutral-angry faces in emotion condition, male-female in gender condition). We will present each face on the screen until participants press the spacebar to indicate that they have updated both internal counters. This response latency for updating is the main dependent variable of the task. The next face appears on the screen after a 200ms inter-stimulus interval. Due to the face-sequence, there are shifts and no shifts in each block of items (e.g. in the emotion condition shifts are angry-neutral and neutral-angry and no shifts are angry-angry and neutral-neutral).[44]

1.1.2.6. Dutch adult reading test (10min)

Dutch version of the national adult reading test. The score is predictive of premorbid intelligence in brain damaged patients and appeared insensitive to brain deterioration in demented or psychotic patients.[45]

1.1.3. Experience sampling method (ESM)

Momentary assessment techniques allow for examination of subtle fluctuations of behaviour and affect over the course of the day, and the prospective nature of the data allows for examination of the temporal association between different observations. The shift to the micro-level of daily life showed how subtle dynamic patterns of moment-to-moment affective experiences and responses to situations constitute the missing link between macro-level risk factors for psychiatric disorders like MDD and future outcomes.[46] Major MDD risk factors interactively impact on reactivity and duration of momentary experiences in everyday life and the latter patterns in turn predict future course of symptoms. Therefore, it is relevant to examine mechanisms at the level of these smallest building blocks. Furthermore, real-life tracking of experiences using ESM might allow for an easy identification of the concrete bits of real-life affective and behavioural patterns which need remediation. Furthermore, ESM can provide real-life validity to experimental and imaging results. This is important, as this clarifies how knowledge of mechanisms connects with real-life intervention targets.[47-49]

The ESM-palmtop ("PsyMate"®) will signal subjects at random moments during the day to answer questions about affect and daily events. Answering the ESM-palmtop questions after each auditory signal ("beep") will take about 30sec. We will program the ESM-palmtop to emit 10 beeps/day at random intervals in each of the ten 90-minutes time blocks between 7:30h and 22:30h, on 6 consecutive days. After each beep, subjects have to fill out the self-assessment on the ESM-palmtop to record current context (activity, persons present, location, physical activity), stress appraisals of this context, and mood. Mood questions include 4 Positive and 5 Negative Affect items.[50] Examples are 'happy' and 'relaxed' for positive affect and 'depressed' and 'irritated' for negative affect. The self-assessments will be rated on 7-point Likert scales (ranging from 1= 'not at all' to 7= 'very'). We will instruct subjects to complete the ESM-palmtop measurements as quickly as possible after the beep. This emphasis helps to minimize retrospective memory distortion. In addition, we included a morning and an evening questionnaire including specific questions regarding sleep and the overall day, respectively. We will instruct participants to fill in these questionnaires on the ESM-palmtop in the morning after they wake up, and in the evening before they go to bed. The questions we included in the ESM-procedure are shown in Supplementary table 1.

Regarding the ESM a standard approach for data cleaning will be used. We will first check for missing data. Second, we will check whether total response time exceeds 15 minutes or whether time between the beep and first response exceeds 15 minutes, which observations will be removed. Third, we will exclude days of measurement when the number of observations was less than five. Fourth, we will exclude subjects when the number of observations is less than 30. These precautions are taken to have enough and valid measurements, necessary for valid statistical approaches. We will thereafter inspect the variables to see whether they contain variation based on the interquartile range.

Because ESM observations are irregularly spaced (due to the random presentation of measurements and missing data) and a positive/negative autocorrelation may exist between the expected absolute successive difference (EASD) and time intervals, we will calculate the mean adjusted absolute successive difference (MAASD) per ESM variable, taking into account an adjustment parameter λ, to capture affective instability.[51] To avoid night time intervals, successive differences will be calculated within days.

Because ESM-data will likely be skewed to the left, we will apply nonparametric independent samples Mann-Whitney U tests when appropriate, to determine significance of differences between the remitted recurrent MDD and healthy control groups.

1.1.4. MRI-scanning

We will minimize side-to-side head movements by fitting foam pads between the volunteers' head and the volume coil. We will obtain scans in the order indicated below; half-way, we will plan a break of >20min. Due to time constraints we had to perform the emotion regulation task (ERT) in the second block of scanning, after the break. It could be questioned whether the brain activation by this ERT might influence the 2nd resting state scan after a mood-induction. If anything, we aimed to have the subjects maximally experience a sad mood after the mood-induction. As the ERT also provided negative pictures, (alternated with positive) it can be expected that the ERT might also have primed people to be more susceptible for the mood induction procedure. As this was a systematic order of scanning in all subjects, we think that if any effect occurred, this would have only primed all subjects systematically to be more vulnerable for the mood induction. Supplementary table 2 describes the experimental designs of all fMRI tasks according to available reporting guidelines,[52] the remaining acquisition parameters are described below.

- Locater scan: a whole brain low resolution 3-dimensional T1-weighted turbo field echo-scan for anatomical overview. Scan duration=53s; number of slices=100; slice orientation=sagittal; field of view (FOV)=250×250×220mm; voxel size 2.23×2.23×2.2mm; acquired matrix=112×112; act. repetition time (TR)=3.1ms; act. echo time (TE) 1.4; flip angle (FA)=8˚; turbo-factor=425.
- Reference scan: to obtain a whole brain sensitivity map for the subsequent SENSitivity Encoding (SENSE) scans. Scan duration=59s; number of slices=100; slice gap=10mm; slice orientation=coronal; FOV=530×530mm; voxel size 5.52×7.07×3mm; acquired matrix=96×75; TR=4ms; TE=0.75; FA=1˚.
- *Structural scan* (6min): a whole brain high resolution 3-dimensional T1-weighted turbo field echo-scan for detailed anatomic information. Scan duration=372s; number of slices=220; slice orientation=transverse; FOV=240×220×188mm; voxel size 1×1×1mm; acquired matrix=240×187; TR=8.3ms; TE=3.8; FA=8˚; number of averages=2; TURBO-factor=154.
- *Resting-state scan*: We will give no specific instructions except that all subjects keep their eyes closed, let their mind wander, lie still and not fall asleep.[53] Because we aim to compare resting-state scans without and with a negative mood induction, we will play neutral or sad music, respectively, the 5min preceding the resting-state scans. We will combine this with a personalized neutral and sad script, respectively, which subjects read on the screen, as described in the main article. We chose not to counterbalance the acquisition of the neutral/mood induced resting state scans, to prevent potential interference of the subsequent fMRI-tasks by mood-state after an initial sad mood-induced resting state scan. This scan will be a field echo (FE) echo-

planar imaging (EPI)-scan with duration=428s; number of slices=37; slice thickness=3mm; act. slice gap=0.3mm; slice orientation=transverse; slice order=ascending; number of dynamics=210; FOV=240×220×122mm; voxel size 3×3×3mm; acquired matrix=80×80; TR=2000ms; TE=28ms; FA=76˚; EPI-factor=43.

- *Reinforcement learning task* (25min): After instructing the participants to arrive thirsty, a Pavlovian-learning paradigm will be used, delivering small amounts (0.2ml) of liquid (sweet apple juice or bitter 3.0M $MgSO_4$) at different probabilities (80-20%) after conditional stimuli. With the changing probabilities of water delivery, temporal difference reward-learning and aversive-learning signals can be calculated which will be used as a regressor of interest in the analyses. This task showed excellent (differential) activations of the reinforcement learning circuitry in depressed subjects versus controls.[54] Of note, the task does not test social stimuli,[55] but rather the persistence of difficulties in temporal difference reward related learning with primary rewards, as this could be a more general and basic persistent dysfunction in recurrent MDD. $MgSO_4$ is clinically used as a laxative, and is not harmful to humans. The 15ml solution used in the experiment will not cause bowel distress. This will be a FEEPI-scan with duration=1693.5s; voxel size 3×3×3mm; EPI-factor=43.
- *Magnetic resonance spectroscopy* (MRS): Glutamate, glutamine and GABA only recently became distinguishable from each other by MRS. We will acquire edited ^{1}H J-difference spectra using a GABA-specific MEGA-PRESS sequence.[56-58] During the odd transients in this sequence, we will apply a 15.64ms sinc-center editing pulse (64 Hz full width at half maximum) at 1.9ppm and 4.6ppm in an interleaved manner to specifically excite GABA and suppress water, respectively. We will acquire these spectra in two voxels, one in the left basal ganglia with scan duration=776s; volume of interest size=30×20×20mm; number of dynamics=384; number of rest slabs =4; number of samples =2048; TR=2000; TE=73ms; FA=90˚; odd frequency=351; even frequency=-351; 2^{nd} order pencil beam-auto shimming; and water suppression. The same in the pgACC, except with scan duration=328s; volume of interest size=25×20×30mm; number of dynamics=160.
- *Diffusion Tensor Imaging* (DTI): measures whole brain fractional anisotropy (FA) and mean diffusivity which can quantify white matter abnormalities.[59] Spin-echo diffusion weighted imaging DTI-scan duration=333.6s; number of slices=60; slice thickness=2mm; slice gap=0mm; slice orientation=transverse; FOV=224×224×120mm; voxel size 2×2×2mm; acquired matrix=112×112; TR=7635ms; TE=88ms; FA=90˚; EPI-factor=59; number of b-factors=2; b-factor order=ascending; max b-factor=1000.
- *Cued Emotional Conflict Task* (CECT,[60] 25min): Participants will be instructed to respond as quickly as possible with two response buttons indicating happy or sad. In an event-related

paradigm, each trial starts with one of two word cues ("actual" or "opposite") presented for 500ms, which instructs participants to respond to the target cue with the identical or opposite valenced button. After the presentation of the cue word, a fixed interval of 2000ms separates the presentation of the cue from the target. The target cue is either a happy or sad face presented in the centre of the screen. This cue-offset period makes it possible to investigate: 1) cue related conflict anticipation; and 2) response related cognitive control following the presentation of the emotional target.

Fourteen faces (7 female and 7 male actors) from the Karolinska Directed Emotional Faces dataset[61] will be used. Each face will be shown in a happy or sad expression (matched for arousal). The assignment of labels to the two response buttons will be counterbalanced across participants. After the CECT, participants will be asked to rate the faces for valence and arousal using 9-point Likert scales (*valence*: 1=unhappy, 5=neutral, 9=happy; *arousal*: 1=calm, 5=intermediate, 9=excited). This will be a FEEPI-scan with voxel size 3×3×3mm; EPI-factor=43. Six runs of 24 trials are separated by a short brake.

- *Emotion regulation task* (ERT, 20min): this will be a modification of the emotion regulation task described earlier.[62 63] The stimulus set will consist of 9 x 4 (sad, happy, fearful, neutral) x 2 (attend, regulate) pictures derived from the International Affective Picture System (IAPS)[64] and http://nl.dreamstime.com; we will match each set for valence, arousal and content. We preselected the IAPS pictures based on IAPS ratings (scale: 1-9) for valence (neutral: 4-6; positive, i.e. happy: >6; negative, i.e. fearful and sad: <4); arousal (neutral: <3; emotional, i.e. happy, fearful, or sad: >6) and furthermore ratings for emotion specificity as assessed by Mikels et al.[64] (scale: 1-9) (>7 for each specific emotion category; neutral: <3 for every emotion). In addition, we will use stock photos from http://nl.dreamstime.com based on emotional content. In total, we selected 110 pairs of pictures, matched for emotional content. To make matching between IAPS and Dreamstime pictures possible, we performed an independent pilot study (N=41 healthy controls). Subjects rated all pictures on valence and arousal [using the same Self-Assessment Manikin (SAM)[65] used for the IAPS database, ranging from unpleasant to pleasant for valence, and from calm to excited for arousal], emotion type [on a scale from 1 (emotion is not elicited at all) till 9 (emotion is elicited very strongly)] and complexity [on a scale from 1 (picture is very easy to interpret) till 9 (picture is very difficult to interpret)]. Based on these ratings, we eventually selected 36 sets of 2 pictures (9 sets for each emotional category). Within each pair, we matched the pictures (one for the attend, one for the regulate condition) for valence, arousal, complexity and emotional content. We will present the pictures in a semi-blocked pseudo-randomized design. Each block will start with the instruction presented in the middle of the screen (4s), followed by 3 successive pictures of the same emotional category (10s each). After each picture,

subjects will indicate the emotional intensity resulting from attending or regulating on a Visual Analogue Scale (VAS). After each block subjects will also rate their performance (i.e. how well they were capable of attending or regulating). We will separate blocks by a fixation cross (4s). We will pseudorandomize and counterbalance the order of stimuli presentation, and instruction within and between subject groups. We expect this task to show – amongst others- negative bold responses when viewing pictures relative to the fixation cross (resting state) in the pregenual anterior cingulate. This will be two FEEPI-scan with max durations=822s; voxel size 3×3×3mm; EPI-factor=43.

1.1.5. Blood measures

1.1.5.1. Fatty acid metabolism

We will wash erythrocytes of venous EDTA blood three times in isotonic saline, count them by routine hemocytometric analysis and freeze them overnight in a BHT (2,6-di-tert-butyl-4-methylphenol)-coated Eppendorf cup. Next, we will transmethylate fifty microliters of the resulting hemolysate in 1ml 3M HCl by incubating for 4hrs at 90°C in the presence of 10nmol internal standard; the methyl ester of 18-methylnonadecanoic acid. After cooling, we will extract the aqueous layer in 2ml hexane, and take this extract to dryness under nitrogen flow and resuspend it in 80µl hexane. Subsequently, we will inject one microliter of this solution into a Hewlett Packard GC 5890 equipped with an Agilent J&W HP-FFAP, 25m, 0.20mm, 0.33µm GC Column, and detect eluting fatty acid methylesters by flame ionization detection. Finally, we will calculate fatty acid concentrations using the known amount of internal standard and express them as pmol/10^6 cells for erythrocytes.[66] [67]

1.1.5.2. Genetics

We will apply polymerase chain reaction (PCR) and HinfI restriction enzyme digestion as described previously.[68] In short, we will isolate DNA from blood using a filter-based method (QIAamp DNA Mini Kit, Qiagen Ltd., United Kingdom). Next, we will design PCR-primers using Primer 3 (available at http://bioinfo.ut.ee/primer3/). Subsequently, we will use a Matrix Assisted Laser Desorption Ionization Time Of Flight (MALDI-TOF) mass spectrometer from Bruker Daltonics. To increase reliability, we will genotype all samples in duplicate. Finally, we will save additional genetic material for future analyses.[14]

1.1.5.3. Blood storage

We will acquire platelet-poor plasma from lithium-heparine, EDTA and citrate blood tubes using the following procedure. First, we will centrifuge tubes for 10min at 2680×g (no brake) at 18°C. Next, we

will pipet plasma of each tube in a separate cryovial®. Subsequently, we will centrifuge the cryovials® for 5min at 14.000×g (no brake) at 18°C. Finally, we will store the platelet-poor plasma in separate micronic vials at -80°C until further analyses.

1.1.6. Salivary measures

We will instruct participants to provide five saliva samples over a (working, if applicable) day before the first study-session (at awakening, 30, 45 and 60min thereafter, followed by a fifth measurement at 22.00h) to diurnally reflect the morning awakening curve and evening HPA-axis activity. In addition, we will collect saliva using a Salivette before and after sad mood induction to investigate HPA-axis response to a psychological personal stressor. While Dickerson & Kemeny[69] indicated that, on average, an emotion induction stressor does not elicit a significant cortisol response, this should be interpreted with caution because of the relatively small numbers of studies that fell in this category as acknowledged by the authors. Moreover, it could be that some studies observed a positive, and others a negative effect, that levels out as no effect overall. In addition, Dickerson & Kemeny excluded studies in which recruitment was based on a physical or psychological diagnosis or a stressful experience (e.g., diabetes, depression, bereavement). This makes it hard to extrapolate their findings to our sample of recurrently depressed patients. Of note, several more recent papers did observe interesting effects of mood on salivary cortisol in recurrent depression (Chopra et al., 2008;[70] Huffziger et al., 2013)[71], which makes this assessment of great interest to our study. We will instruct subjects not to eat, smoke, drink tea or coffee or brush their teeth within 15min before sampling,[10 72 73] to write down exact sampling day and time, and to keep samples refrigerated before bringing them back to us at the first study-session; we will store them (-20°C) until analysis.

1.1.7. Waist circumference

Increased waist circumference reflects abdominal obesity, which is a metabolic syndrome criterion.[74-77] Abdominal obesity is closely related to insulin resistance and metabolic dysregulation, and a strong risk factor for development of diabetes type II and cardiovascular disease.[78] We will measure waist circumference at the vertical middle between the lowest palpable rib and upper part of the ilium. We will use a solid, nonexpendable, measuring tape, which we will apply with light pressure (but without squeezing underlying tissues) horizontally around the waist. We will instruct subjects to stand with their feet close together, arms next to their body, and their bodyweight equally distributed. We will instruct subjects to take of thick clothing, and perform the actual measurement at the end of a normal expiration.

1.2. Power analyses

Power analyses were performed using G*Power 3.9.1.2 and package PowerSurvEpi in R.[79-83] In the cross-sectional compararison between patients (n = 60) and controls (n =40), with 80% power, α=.05 and 5 predictors in total we can detect small effects (effect size f^2 = 0.0994846). For the prospective analyses, Cox proportional hazard regression with 60 patients of which 2×±20 eligble for a second scan , a correlated covariate of interest and a moderate effect size results in a power of >80% with α=.05.

Of note, in the initially registered trial protocotol we proposed to include 50 patients and 50 controls. However, based on data from our previous studies that was analysed since then[10 14 66 84-86] we amended this aspect in our protocol to 60 patients and 40 controls. While yielding an identical total number of subjects, this provides a more optimized balance between the contrast patients vs. controls on the one hand, and the prospective analyses in the patients on the other. Analyses of our previous studies show that there exist rather large effects in the differences between patients and controls, and relatively smaller effects in the prospective associations. By changing the patient:control ratio to 60:40, we lose only little power in the cross-sectional analyses (a little less optimal distribution but equal total number), but gain additional prospective power. In addition, because the estimates in the control population are expected to be more homogeneous than in the patient population, also in the cross-sectional analyses the decrease in sample size of the control population is expected to result in smaller loss of power than the gain in power resulting from the equal increase of the patient sample size.

REFERENCES

1. Hamilton M. A rating scale for depression. Journal of neurology, neurosurgery, and psychiatry 1960;**23**:56-62.
2. John Rush A, Giles DE, Schlesser MA, et al. The inventory for depressive symptomatology (IDS): Preliminary findings. Psychiatry Research 1986;**18**(1):65-87.
3. Oliver JM, Baumgart EP. The Dysfunctional Attitude Scale - Psychometric Properties and Relation to Depression in an Unselected Adult-Population. Cognit Ther Res 1985;**9**(2):161-67.
4. Rush AJ, Gullion CM, Basco MR, et al. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996;**26**(3):477-86.
5. Van der Does W. Cognitive reactivity to sad mood: structure and validity of a new measure. Behav Res Ther 2002;**40**(1):105-20.
6. Moulds ML, Kandris E, Williams AD, et al. An investigation of the relationship between cognitive reactivity and rumination. Behav Ther 2008;**39**(1):65-71.
7. Murray G, Rawlings D, Allen NB, et al. NEO five-factor inventory scores: Psychometric properties in a community sample. Meas Eval Couns Dev 2003;**36**(3):140-49.
8. McCrae RR, Costa PT. A contemplated revision of the NEO Five-Factor Inventory. Pers Individ Dif 2004;**36**(3):587-96.
9. Vingerhoets AJJM, Jeninga AJ, Menges LJ. The Measurement of Daily Hassles and Chronic Stressors - the Development of the Everyday Problem Checklist (Epcl, Dutch - Apl). Gedrag Gezond 1989;**17**(1):10-17.
10. Lok A, Mocking RJ, Ruhé HG, et al. Longitudinal hypothalamic-pituitary-adrenal axis trait and state effects in recurrent depression. Psychoneuroendocrinology 2012;**37**(7):892-902.
11. Turner H, Bryant-Waugh R, Peveler R, et al. A psychometric evaluation of an English version of the Utrecht Coping List. Eur Eat Disord Rev 2012;**20**(4):339-42.
12. Kraaij V, de Wilde EJ. Negative life events and depressive symptoms in the elderly: a life span perspective. Aging & mental health 2001;**5**(1):84-91.
13. Thombs BD, Bernstein DP, Lobbestael J, et al. A validation study of the Dutch Childhood Trauma Questionnaire-Short Form: factor structure, reliability, and known-groups validity. Child Abuse Negl 2009;**33**(8):518-23.
14. Lok A, Bockting CL, Koeter MW, et al. Interaction between the MTHFR C677T polymorphism and traumatic childhood events predicts depression. Translational psychiatry 2013;**3**:e288.
15. Bernstein DP, Ahluvalia T, Pogge D, et al. Validity of the Childhood Trauma Questionnaire in an adolescent psychiatric population. J Am Acad Child Adolesc Psychiatry 1997;**36**(3):340-8.
16. Bauman A, Ainsworth BE, Bull F, et al. Progress and Pitfalls in the Use of the International Physical Activity Questionnaire (IPAQ) for Adult Physical Activity Surveillance. J Phys Act Health 2009;**6**:S5-S8.
17. Lee PH, Macfarlane DJ, Lam TH, et al. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review, 2011:115-15.
18. Vandelanotte C, De Bourdeaudhuij I, Philippaerts R, et al. Reliability and validity of a computerized and Dutch version of the International Physical Activity Questionnaire (IPAQ). 2005.
19. Faulkner G, Cohn T, Remington G. Validation of a physical activity assessment tool for individuals with schizophrenia. Schizophr Res 2006;**82**(2-3):225-31.
20. Beukers MH, Dekker LH, de Boer EJ, et al. Development of the HELIUS food frequency questionnaires: ethnic-specific questionnaires to assess the diet of a multiethnic population in The Netherlands. Eur J Clin Nutr 2014.
21. Dekker LH, Snijder MB, Beukers MH, et al. A prospective cohort study of dietary patterns of non-western migrants in the Netherlands in relation to risk factors for cardiovascular diseases: HELIUS-Dietary Patterns. BMC Public Health 2011;**11**:441.

22. Levine DW, Kripke DF, Kaplan RM, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. Psychol Assess 2003;**15**(2):137-48.
23. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971;**9**(1):97-113.
24. Veale JF. Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. Laterality 2014;**19**(2):164-77.
25. Snaith RP, Hamilton M, Morley S, et al. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br J Psychiatry 1995;**167**(1):99-103.
26. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. J Pers Soc Psychol 1991;**61**(1):115-21.
27. Schoofs H, Hermans D, Raes F. Brooding and Reflection as Subtypes of Rumination: Evidence from Confirmatory Factor Analysis in Nonclinical Samples using the Dutch Ruminative Response Scale. J Psychopathol Behav 2010;**32**(4):609-17.
28. van Rijsbergen GD, Kok GD, Elgersma HJ, et al. Personality and cognitive vulnerability in remitted recurrently depressed patients. J Affect Disord 2015;**173**:97-104.
29. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011;**63 Suppl 11**:S467-72.
30. Wardenaar KJ, van Veen T, Giltay EJ, et al. Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). Psychiatry Res 2010;**179**(1):101-6.
31. Parker G, Hadzi-Pavlovic D, Boyce P, et al. Classifying depression by mental state signs. Br J Psychiatry 1990;**157**:55-65.
32. Parker G. Is the diagnosis of melancholia important in shaping clinical management? Curr Opin Psychiatr 2007;**20**(3):197-201.
33. Parker G. Defining melancholia: the primacy of psychomotor disturbance. Acta Psychiatr Scand Suppl 2007;**115**(433):21-30.
34. Bennabi D, Vandel P, Papaxanthis C, et al. Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiologic, and therapeutic implications. Biomed Res Int 2013;**2013**:158746.
35. Rhebergen D, Arts DL, Comijs H, et al. Psychometric properties of the dutch version of the core measure of melancholia. J Affect Disord 2012;**142**(1-3):343-6.
36. Dantchev N, Widlocher DJ. The measurement of retardation in depression. The Journal of clinical psychiatry 1998;**59 Suppl 14**:19-25.
37. Schrijvers D, de Bruijn ER, Maas Y, et al. Action monitoring in major depressive disorder with psychomotor retardation. Cortex 2008;**44**(5):569-79.
38. Posner MI, Snyder CR, Davidson BJ. Attention and the detection of signals. J Exp Psychol 1980;**109**(2):160-74.
39. Leyman L, De Raedt R, Schacht R, et al. Attentional biases for angry faces in unipolar depression. Psychol Med 2007;**37**(3):393-402.
40. Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. Br J Psychiatry 2009;**195**(2):102-8.
41. Mocking RJ, Patrick Pflanz C, Pringle A, et al. Effects of short-term varenicline administration on emotional and cognitive processing in healthy, non-smoking adults: a randomized, double-blind, study. Neuropsychopharmacology 2013;**38**(3):476-84.
42. Harmer CJ, O'Sullivan U, Favaron E, et al. Effect of acute antidepressant administration on negative affective bias in depressed patients. Am J Psychiatry 2009;**166**(10):1178-84.
43. Chambers R, Lo BCY, Allen NB. The impact of intensive mindfulness training on attentional control, cognitive style, and affect. Cognit Ther Res 2008;**32**(3):303-22.
44. De Lissnyder E, Koster EH, De Raedt R. Emotional interference in working memory is related to rumination. Cognit Ther Res 2012;**36**(4):348-57.

45. Schmand B, Bakker D, Saan R, et al. [The Dutch Reading Test for Adults: a measure of premorbid intelligence level]. Tijdschr Gerontol Geriatr 1991;**22**(1):15-9.
46. Wichers M, Peeters F, Geschwind N, et al. Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. Journal of Affective Disorders 2010;**124**(1-2):191-95.
47. Wichers M, Lothmann C, Simons CJP, et al. The dynamic interplay between negative and positive emotions in daily life predicts response to treatment in depression: A momentary assessment study. Br J Clin Psychol 2012;**51**:206-22.
48. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;**4**:1-32.
49. Myin-Germeys I, Oorschot M, Collip D, et al. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009;**39**(9):1533-47.
50. Crawford JR, Henry JD. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004;**43**(Pt 3):245-65.
51. Jahng S, Wood PK, Trull TJ. Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. Psychol Methods 2008;**13**(4):354-75.
52. Guo Q, Parlar M, Truong W, et al. The reporting of observational clinical functional magnetic resonance imaging studies: a systematic review. PloS one 2014;**9**(4):e94412.
53. Greicius MD, Flores BH, Menon V, et al. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biol Psychiatry 2007;**62**(5):429-37.
54. Kumar P, Waiter G, Ahearn T, et al. Abnormal temporal difference reward-learning signals in major depression. Brain 2008;**131**(Pt 8):2084-93.
55. Davey CG, Allen NB, Harrison BJ, et al. Being liked activates primary reward and midline self-related brain regions. Hum Brain Mapp 2010;**31**(4):660-8.
56. Waddell KW, Avison MJ, Joers JM, et al. A practical guide to robust detection of GABA in human brain by J-difference spectroscopy at 3 T using a standard volume coil. Magn Reson Imaging 2007;**25**(7):1032-8.
57. van Loon Anouk M, Knapen T, Scholte HS, et al. GABA Shapes the Dynamics of Bistable Perception. Curr Biol 2013;**23**(9):823-27.
58. Waddell KW, Zanjanipour P, Pradhan S, et al. Anterior cingulate and cerebellar GABA and Glu correlations measured by 1H J-difference spectroscopy. Magn Reson Imaging 2011;**29**(1):19-24.
59. Taylor WD, Hsu E, Krishnan KR, et al. Diffusion tensor imaging: background, potential, and utility in psychiatric research. Biol Psychiatry 2004;**55**(3):201-7.
60. Vanderhasselt MA, Baeken C, Van Schuerbeek P, et al. How brooding minds inhibit negative material: An event-related fMRI study. Brain Cogn 2013;**81**(3):352-59.
61. Goeleven E, De Raedt R, Leyman L, et al. The Karolinska Directed Emotional Faces: A validation study. Cognition Emotion 2008;**22**(6):1094-118.
62. Levesque J, Eugene F, Joanette Y, et al. Neural circuitry underlying voluntary suppression of sadness. Biol Psychiatry 2003;**53**(6):502-10.
63. Johnstone T, van Reekum CM, Urry HL, et al. Failure to regulate: counterproductive recruitment of top-down prefrontal-subcortical circuitry in major depression. J Neurosci 2007;**27**(33):8877-84.
64. Mikels JA, Fredrickson BL, Larkin GR, et al. Emotional category data on images from the International Affective Picture System. Behav Res Methods 2005;**37**(4):626-30.
65. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. J Behav Ther Exp Psychiatry 1994;**25**(1):49-59.
66. Assies J, Pouwer F, Lok A, et al. Plasma and erythrocyte fatty acid patterns in patients with recurrent depression: a matched case-control study. PloS one 2010;**5**(5):e10635.

67. Mocking RJ, Assies J, Lok A, et al. Statistical methodological issues in handling of fatty acid data: percentage or concentration, imputation and indices. Lipids 2012;**47**(5):541-7.
68. Frosst P, Blom HJ, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 1995;**10**(1):111-3.
69. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004;**130**(3):355-91.
70. Chopra KK, Segal ZV, Buis T, et al. Investigating associations between cortisol and cognitive reactivity to sad mood provocation and the prediction of relapse in remitted major depression. Asian J Psychiatr 2008;**1**(2):33-6.
71. Huffziger S, Ebner-Priemer U, Zamoscik V, et al. Effects of mood and rumination on cortisol levels in daily life: an ambulatory assessment study in remitted depressed patients and healthy controls. Psychoneuroendocrinology 2013;**38**(10):2258-67.
72. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. Psychoneuroendocrinology 1994;**19**(4):313-33.
73. Vreeburg SA, Kruijtzer BP, van Pelt J, et al. Associations between sociodemographic, sampling and health factors and various salivary cortisol indicators in a large sample without psychopathology. Psychoneuroendocrinology 2009;**34**(8):1109-20.
74. Ohaeri JU, Akanji AO. Metabolic syndrome in severe mental disorders. Metab Syndr Relat Disord 2011;**9**(2):91-8.
75. Kahl KG, Greggersen W, Schweiger U, et al. Prevalence of the metabolic syndrome in unipolar major depression. Eur Arch Psychiatry Clin Neurosci 2012;**262**(4):313-20.
76. Oda E. Metabolic syndrome: its history, mechanisms, and limitations. Acta Diabetol 2012;**49**(2):89-95.
77. Fisman EZ, Tenenbaum A. The metabolic syndrome entanglement: Cutting the Gordian knot. Cardiol J 2014;**21**(1):1-5.
78. Assies J, Mocking RJ, Lok A, et al. Effects of oxidative stress on fatty acid- and one-carbon-metabolism in psychiatric and cardiovascular disease comorbidity. Acta psychiatrica Scandinavica 2014;**130**(3):163-80.
79. Faul F, Erdfelder E, Buchner A, et al. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;**41**(4):1149-60.
80. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics 1983;**39**(2):499-503.
81. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. Stat Med 2004;**23**(21):3263-74.
82. Power and sample size calculation for survival analysis of epidemiological studies [program], 2015.
83. R: A language and environment for statistical computing. R Foundation for Statistical Computing [program]. Vienna, Austria, 2013.
84. Bockting CL, Mocking RJ, Lok A, et al. Therapygenetics: the 5HTTLPR as a biomarker for response to psychological therapy? MolPsychiatry 2013;**18**(7):744-45.
85. Lok A, Assies J, Koeter MW, et al. Sustained medically unexplained physical symptoms in euthymic patients with recurrent depression: predictive value for recurrence and associations with omega 3- and 6 fatty acids and 5-HTTLPR? J Affect Disord 2012;**136**(3):604-11.
86. Lok A, Mocking RJ, Assies J, et al. The one-carbon-cycle and methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism in recurrent major depressive disorder; influence of antidepressant use and depressive state? J Affect Disord 2014;**166**:115-23.