

University of Groningen

## Increased affective reactivity among depressed individuals can be explained by floor effects

von Klipstein, Lino; Servaas, Michelle N; Lamers, Femke; Schoevers, Robert A; Wardenaar, Klaas J; Riese, Harriëtte

*Published in:*  
Journal of Affective Disorders

*DOI:*  
[10.1016/j.jad.2023.04.118](https://doi.org/10.1016/j.jad.2023.04.118)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2023

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

von Klipstein, L., Servaas, M. N., Lamers, F., Schoevers, R. A., Wardenaar, K. J., & Riese, H. (2023). Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study. *Journal of Affective Disorders*, 334, 370-381. <https://doi.org/10.1016/j.jad.2023.04.118>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



## Research paper

# Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study

Lino von Klipstein<sup>a,\*</sup>, Michelle N. Servaas<sup>a</sup>, Femke Lamers<sup>b</sup>, Robert A. Schoevers<sup>a</sup>, Klaas J. Wardenaar<sup>a</sup>, Harriëtte Riese<sup>a</sup>

<sup>a</sup> University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), the Netherlands

<sup>b</sup> Department of Psychiatry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health research institute, Amsterdam, the Netherlands



## ARTICLE INFO

## Keywords:

Emotional reactivity  
Stress reactivity  
Major Depressive Disorder  
Ecological momentary assessment  
Bounded measurement

## ABSTRACT

Experience sampling studies into daily-life affective reactivity indicate that depressed individuals react more strongly to both positive and negative stimuli than non-depressed individuals, particularly on negative affect (NA). Given the different mean levels of both positive affect (PA) and NA between patients and controls, such findings may be influenced by floor/ceiling effects, leading to violations of the normality and homoscedasticity assumptions underlying the used statistical models. Affect distributions in prior studies suggest that this may have particularly influenced NA-reactivity findings. Here, we investigated the influence of floor/ceiling effects on the observed PA- and NA-reactivity to both positive and negative events. Data came from 346 depressed, non-depressed, and remitted participants from the Netherlands Study of Depression and Anxiety (NESDA). In PA-reactivity analyses, no floor/ceiling effects and assumption violations were observed, and PA-reactivity to positive events, but not negative events, was significantly increased in the depressed and remitted groups versus the non-depressed group. However, NA-scores exhibited a floor effect in the non-depressed group and naively estimated models violated model assumptions. When these violations were accounted for in subsequent analyses, group differences in NA-reactivity that had been present in the naive models were no longer observed. In conclusion, we found increased PA-reactivity to positive events but no evidence of increased NA-reactivity in depressed individuals when accounting for violations of assumptions. The results indicate that affective-reactivity results are very sensitive to modeling choices and that previously observed increased NA-reactivity in depressed individuals may (partially) reflect unaddressed assumption violations resulting from floor effects in NA.

## 1. Introduction

Over the past decades the experience sampling method (ESM) has become an indispensable tool for the study of affect in the daily lives of individuals (Houben et al., 2015). Although single cross-sectional measurements can be used to answer many different research questions, intensive repeated measurements are required to catch fluctuations in affect over time and patterns herein. Nowadays, ESM typically involves participants filling out short questionnaires on their momentary positive affect (PA) and negative affect (NA) on their smartphone multiple times per day for weeks or even months. The ESM approach is widely used in depression research to investigate various metrics of

dynamic change of affect over time, such as affective inertia, affective instability, and affective reactivity (e.g., Nelson et al., 2020; Telford et al., 2012; Thompson et al., 2012). These dynamic aspects are currently not as well understood as depression's characteristic negative emotional tone (referred to as mood; Rottenberg, 2017), fueling continued research interest.

Here, we focus on affective reactivity, which pertains to how strong an individual's affect reacts to stimuli (e.g., negative events) in their daily lives. ESM studies have typically operationalized affective reactivity as the association between momentary NA or PA and the co-occurring presence of momentary contextual stimuli, with some variation in the precise details (e.g., differences in ESM designs and in

\* Corresponding author at: University of Groningen, University Medical Center Groningen, Groningen, Department of Psychiatry, PO Box 30.001 (CC72), 9700 RB Groningen, the Netherlands.

E-mail address: [p.l.von.klipstein@umcg.nl](mailto:p.l.von.klipstein@umcg.nl) (L. von Klipstein).

<https://doi.org/10.1016/j.jad.2023.04.118>

Received 15 December 2022; Received in revised form 11 April 2023; Accepted 29 April 2023

Available online 6 May 2023

0165-0327/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

modeling choices, such as the use of person-mean-centering and controlling for lagged affect). Affective reactivity is unique among affect-dynamic metrics, as it directly relates affect to the environment. Further, it is especially interesting in the depression context, as the altered interaction of affect and environment is both a defining feature of depression (i.e., reduced ability to experience pleasure; [American Psychiatric Association, 2013](#)) and is theorized to be a major etiological and maintenance factor (i.e., behavioral activation model of depression; [Jacobson et al., 2001](#)).

Although results from ESM research on affective reactivity have been insightful, they may have been influenced by methodological issues that have recently been raised in the context of other affect dynamics. [Mestdagh et al. \(2018\)](#) pointed out that if the mean affect level in ESM measurements lies closer to the extreme ends of the measurement scale (e.g., the NA total score), this restricts the possibility for the affect measure to show fluctuations, which suggests that the means and variances of ESM affect measures are confounded. Moreover, when means are extremely low or high (i.e., exhibiting a floor/ceiling effect in the distribution of affect), this can lead to violations of the assumptions of commonly applied statistical regression models (e.g., multivariate normality, homoscedasticity). If left unaddressed, such violations can potentially lead researchers to draw faulty conclusions based on their data ([Ernst and Albers, 2017](#)). [Terluin et al. \(2016\)](#), for example, showed that group differences in associations between affective states were overestimated when violations of regression-model assumptions were not accounted for. Some recent evidence has also shown that various metrics of affect dynamics no longer show an association with depressive symptoms if one accounts for differences in mean affect levels by including the latter as a covariate ([Dejonckheere et al., 2019](#)), though there have also been findings that remained robust after controlling for mean levels ([Schoevers et al., 2020](#)). It has not yet been investigated to what extent these issues influence results on affective reactivity. This is likely because affective reactivity is modelled differently than other metrics of affect dynamics, i.e., using affect as the outcome variable, necessitating a different type of solution. To address this knowledge gap, we investigated the potential role of floor effects in ESM research of affective reactivity in depression in this study.

Floor/ceiling effects occur when a large proportion of scores on a variable fall close to the lower/higher bounds of its scale, leading to data that are characterized by skewed distributions and limited variance (e.g., [Šimkovic and Träuble, 2019](#)). Such distribution characteristics should be acknowledged and addressed in order to avoid violations of the assumptions underlying the most commonly used statistical models. As we will outline below, there are indications of floor effects in the NA distributions in non-depressed control groups in ESM research, which is highly relevant when analyzing reactivity of NA and comparing it between depressed and non-depressed groups. Floor effects in distributions of affect scores likely result from at least two (related) aspects. First, on average, non-depressed individuals experience little NA in their daily lives, consistent with the absence of depression. Second, the used measurement scales have bounded ranges, causing non-depressed individuals to frequently ‘bottom out’ on the NA scale. Such floor effects can lead to violations of statistical assumptions in linear multilevel regression models, which are commonly used in ESM affective reactivity research. More specifically, the assumptions of normally distributed residuals, heteroscedasticity of residuals, and linearity may all be violated. Also, floor effects restrict the detectable longitudinal change near the lower bound of the scale as a value at the lower bound cannot decrease any further (e.g., [Fries et al., 2014](#); [Šimkovic and Träuble, 2019](#); [Ward et al., 2014](#)). Because of these modeling issues, floor effects in NA may bias any observed associations with other variables if the model violations are not accounted for. One association that is likely to be affected by such bias is affective reactivity, which has been operationalized as the association between momentary affect and the presence of a daily-life stimulus. If one group of participants exhibits a floor effect in the affect scores (non-depressed individuals), whereas the other

group (depressed individuals) does not, model assumptions of normally distributed residuals (due to extreme skewness in non-depressed individuals) and homoscedasticity (due to restricted variance in non-depressed individuals) are all likely to be violated.

When these violations are not accounted for in the modeling process, interpretation of linear-mixed-model-estimated group differences in affective reactivity becomes problematic. It is recommended that potential model violations are always checked by inspection of the residual distributions, for example using Q-Q plots and residual plots. In case of violations, these should ideally be accounted for statistically before making any model-based inferences. Depending on the specifics of the dataset and the nature of the assumption violations, options include transformation of the outcome (e.g., Log transformation), using other types of regression models within a generalized linear modeling framework (e.g., with a gamma, inverse-gaussian, or Poisson distribution), using non-parametric models, and others (textbooks on these topics include [Faraway, 2016](#); [Gelman and Hill, 2007](#); [Hoffmann, 2021](#); [Kutner et al., 2005](#); [Montgomery et al., 2021](#)). Although widely advocated and documented in the general statistical literature, these recommendations are not always followed ([Ernst and Albers, 2017](#)).

To evaluate if a floor effect in the NA distribution among non-depressed individuals may indeed have influenced findings of increased affective reactivity in depressed versus non-depressed persons, we can take a closer look at prior ESM studies of affective reactivity; more specifically i) whether there was evidence of a floor effect in non-depressed subjects’ NA-score distributions and ii) whether the effect of this floor effect on model assumptions/results was checked and accounted for. To investigate the former, we collected and evaluated the descriptive sample statistics from prior studies (see Table S1 in online supplemental material). In six out of seven studies reporting this information, the mean of the pooled NA values lay close to the lower bound of the scoring range in the non-depressed groups, whereas the NA means were higher in the depressed groups ([Khazanov et al., 2019](#); [Myin-Germeys et al., 2003](#); [Peeters et al., 2003](#); [Thompson et al., 2012](#); [van der Stouwe et al., 2019](#); [van Winkel et al., 2015](#)). Furthermore, standard deviations of NA in non-depressed groups were half or less of the standard deviations of NA in depressed groups in these studies. In the same studies, there was no indication of skewness in the NA-score distributions. NA floor effects may have impacted estimates of group differences in NA reactivity between depressed and non-depressed individuals, most likely in the direction of overestimating them because the restricted NA variation in non-depressed individuals is likely to lead to underestimations of associations involving NA in this group, including NA reactivity. This hypothesis also aligns with the pattern of empirical findings. [Table 1](#) gives an overview of observed group differences between Major Depressive Disorder (MDD) patients and non-depressed individuals in ESM studies on affective reactivity. Most studies found group differences in NA reactivity to various positive and negative stimuli, with depressed individuals showing more NA reactivity than non-depressed individuals, but they found little or no group differences in PA reactivity. Note that increased NA reactivity to positive stimuli refers to stronger *reductions* in NA (i.e., mood-brightening effect in NA), whereas increased NA reactivity to negative stimuli refers to stronger *increases* in NA. ESM research thereby shows increased NA reactivity in both directions in individuals with a depression diagnosis. The question is whether these findings (partially) result from the floor effects in the NA distributions of non-depressed groups, or represent true group differences. While the models employed in previous ESM studies on this subject listed in [Table 1](#) take into account differences in mean affect by estimating random intercepts, no information was provided about assumption checks, which could have revealed assumption violations and the necessity to account for this in the used statistical models.

In the present study we aimed to investigate whether ESM findings on affective reactivity in depression can be replicated when accounting for the assumptions of conditional non-normality and heteroscedasticity, which result from a possible floor effect in non-depressed

**Table 1**

Overview of significant group differences between individuals with an MDD diagnosis and a comparison group in ESM studies on affective reactivity organized by stimulus and response domain.

| Stimulus          | PA-reactivity (in the stimulus-congruent direction) |              |                              | NA-reactivity (in the stimulus-congruent direction) |              |                          |
|-------------------|---|--------------|------------------------------|---|--------------|--------------------------|
|                   | Higher in MDD                                       | Lower in MDD | No differences               | Higher in MDD                                       | Lower in MDD | No differences           |
| Positive event    | 3, 8  |              | 1, 2, 3 <sup>a</sup> , 5, 9  | 1, 3, 6, 7 <sup>b</sup> , 8, 9                      |              | 2, 3 <sup>a</sup> , 5    |
| Positive activity |   |              | 2, 5                         | 5   |              | 2                        |
| Negative event    |   | 8            | 1, 5, 9, 11, 11 <sup>c</sup> | 11  | 8            | 1, 5, 9, 11 <sup>c</sup> |
| Unpleasantness    |   |              |                              | 10  |              |                          |
| Uncontrollability |   |              |                              |   |              | 10                       |
| Unpredictability  |   |              |                              |   |              | 10                       |
| Negative activity |   |              | 4, 5, 11, 11 <sup>c</sup>    | 4, 11   |              | 5, 11 <sup>c</sup>       |
| Social stress     |   |              | 4, 11, 11 <sup>c</sup>       | 4, 11, 11 <sup>c</sup>                              |              |                          |

Note. Studies included here meet the following criteria: (1) investigating a momentary affective reaction to a momentary stimulus, each measured through ESM multiple times per day, (2) investigating relationship of affective reactivity to current depression diagnosis. If not otherwise indicated, comparisons are between an MDD group and a healthy control group. The direction of reactivity is determined by the stimulus, for example NA-reactivity to positive events refers to the reductions in NA, whereas NA-reactivity to negative events refers to increases in NA. In Table S1 more detailed information is provided on the studies included here. Note that results from the same study cannot be considered independent. 1 = [Bylsma et al. \(2011\)](#); 2 = [Heininga et al. \(2019\)](#); 3 = [Khazanov et al. \(2019\)](#); 4 = [Myin-Germeys et al. \(2003\)](#); 5 = [Nelson et al. \(2020\)](#); 6 = [Panaite et al. \(2018\)](#); 7 = [Panaite et al. \(2019\)](#); 8 = [Peeters et al. \(2003\)](#); 9 = [Thompson et al. \(2012\)](#); 10 = [van der Stouwe et al. \(2019\)](#); 11 = [van Winkel et al. \(2015\)](#).

<sup>a</sup> Comparison between MDD and Generalized Anxiety Disorder.

<sup>b</sup> Continuous depression scale including clinical levels.

<sup>c</sup> Comparison between remitted MDD and healthy controls.

individuals' NA scoring distributions. To this end, we used a large sample to investigate differences in momentary affective reactivity (in both PA and NA) to positive and negative momentary events between three groups: individuals with a current depression diagnosis, individuals currently in remission from depression, and individuals that never suffered from depression. We tested the following four hypotheses. (i) NA displays a strongly skewed distribution, indicative of a floor effect in the non-depressed group but not in the group with a current depression diagnosis, whereas PA displays neither floor nor ceiling effects in any group. (ii) When we follow a naive analysis strategy analog to prior ESM affective reactivity studies, the results will show increased reactivity in the currently depressed versus non-depressed group for positive and negative momentary events in NA, but not in PA. (iii) In this naive analysis we will find that assumptions of normally distributed residuals and homogeneity of residual variances are violated in analyses of NA reactivity (iv) When we use a modeling strategy that is better specified given the NA distribution, results will no longer show differences in NA reactivity between the currently depressed and non-depressed groups. As we have data of a unique group of individuals remitted from a depression, who have remained almost unexplored in ESM studies thus far (with one small-sample exception; [van Winkel et al., 2015](#)), we investigated how this group compares to the currently depressed and non-depressed groups on PA and NA reactivity. These comparisons are also expected to be relevant from a methodological perspective, since remitted individuals are less likely to exhibit similarly skewed NA distributions as non-depressed individuals ([Thompson et al., 2021](#)).

Our group previously published an article on group differences in affective instability, a different affect-dynamics metric, using the same dataset ([Schoevers et al., 2020](#)). Affective instability pertains to the amount of moment-to-moment variability in affect, irrespective of environmental stimuli. We had prior knowledge about the distributions of PA and NA in our sample, so that tests on our first hypothesis cannot be considered confirmatory.

## 2. Methods

### 2.1. Sample

The sample consists of participants from the Ecological Momentary Assessment & Actigraphy sub-study (NESDA-EMAA) of the Netherlands

Study of Depression and Anxiety (NESDA). NESDA is an ongoing longitudinal cohort study to examine the long-term course of depressive and anxiety disorders in different health care settings and phases of illness. Detailed information about NESDA is provided elsewhere ([Penninx et al., 2008](#); [Penninx et al., 2021](#)). After the baseline assessment (2004–2007;  $N = 2981$ ), NESDA conducted five follow-up assessments, 1, 2, 4, 6, and 9 years after baseline, including a psychiatric diagnostic interview (details below). At the nine-year follow-up assessment (2014–2017;  $N = 1776$ ), a subset of participants as well as siblings of participants were approached and ultimately included in the NESDA-EMAA study ( $N = 384$ ; for details see [Schoevers et al., 2020](#)). In the current study, to ensure independence of observations between subjects and groups 29 individuals were excluded because they were siblings of other participants in the study. Furthermore, 8 individuals were excluded because they responded to <50 % of ESM assessments, and 1 individual was excluded because their ESM affect measurements showed no variation. This resulted in a sample of  $N = 346$ . The sample was divided into three groups based on their diagnostic history and current diagnostic status of depression assessed in the NESDA assessment waves: i) a depression group (D) with a diagnosis of MDD or Dysthymic Disorder (DD) within the six months before the nine-year follow-up assessment ( $n = 61$ ), ii) a remitted group (rD) with a lifetime diagnosis of MDD or DD, but no diagnosis in the six months prior to their nine-year follow-up assessment ( $n = 189$ ), and iii) a non-depressed group with no lifetime history of MDD or DD (noD,  $n = 96$ ). The NESDA study was approved by the VUmc ethical committee (reference number 2003/183) and all participants gave informed consent prior to enrollment. Data is not publicly available, as true anonymity cannot be guaranteed given the type of data. However, NESDA is committed to accessibility and access can be requested with the NESDA consortium.<sup>1</sup>

### 2.2. ESM assessment

Participants filled out ESM questionnaires for 2 weeks, 5 times per day. Questionnaires were sent at 3-hour intervals via text message to a participant's smartphone and were administered online after opening a secured link in the browser. Data were gathered via the secured server system RoQua ([Sytema and van der Krieke, 2013](#)). Participants were

<sup>1</sup> More information under [www.nesda.nl/nesda-english/](http://www.nesda.nl/nesda-english/).

instructed to complete the questionnaires as soon as possible after receiving the text message (beep), preferably within 15 min, but at least within 60 min. If they had not fill in the diary, they received a text-message reminder after 30 min. After finishing the ESM monitoring, participants received a €20 reimbursement and a personalized report on their ESM data.

## 2.3. Measures

### 2.3.1. Depression diagnosis

Similar to previous waves, at the nine-year follow-up, DSM-IV diagnoses of depressive disorders (MDD and DD) and anxiety disorders (social anxiety disorder, panic disorder with and without agoraphobia, agoraphobia and generalized anxiety disorder) were established with the Composite International Diagnostic Interview (CIDI, version 2.1; Wittchen, 1994). Trained clinical research staff conducted the interviews.

### 2.3.2. Depression and anxiety self-report questionnaires

Severity of depressive symptoms was measured with the 30-item Inventory of Depressive Symptomatology Self Report (IDS-SR; Rush et al., 1996). Severity of anxiety symptoms was measured with the Beck Anxiety Inventory (BAI; Beck et al., 1988; Muntingh et al., 2011). The IDS-SR and BAI are a self-report instrument that served to provide more descriptive information on the present sample.

### 2.3.3. ESM measures

Momentary affect and events were assessed as part of an ESM questionnaire with up to 31 items (the complete list of items is publicly available in the ESM Item Repository; [www.esmitemrepository.com](http://www.esmitemrepository.com)). Affect items covered emotional adjectives on the (positive/negative) valence and (high/low) arousal dimensions (Watson and Tellegen, 1985). They were rated on a 7-point Likert scale ranging from '1 = not at all' to '7 = very much'. A PA scale was calculated by averaging PA items (items: at this moment I feel satisfied, relaxed, cheerful, energetic, enthusiastic, calm) and a NA scale was calculated by averaging NA items (items: at this moment I feel upset, irritated, listless/apathic, down, nervous, bored, anxious). To make the minimal value zero, one point was subtracted from PA and NA scores, resulting in affect variables with range 0–6. This allowed us the use of models aimed at dealing with zero-inflation in later analysis stages. Note that the original scale (1–7) was used in analyses that required only non-zero/non-negative values (e.g., Log transformation). Events were assessed by asking participants 'Did you have daily (un)pleasant experiences since you filled out the previous assessment?', where they could respond 'yes, something pleasant', 'yes, something unpleasant', 'yes, both something pleasant and unpleasant', and 'none'. Answers were recoded into two dichotomous (1 - yes, 0 - no) variables for positive and negative momentary events for the current study.

## 2.4. Statistical analysis

### 2.4.1. Data preprocessing

Preprocessing was performed largely in accordance with Schoevers et al. (2020) and led to the exclusion of 38 participants from the analyses. Details are given in S2 of the online supplemental material. All analyses were conducted in R software (version 4.0.3; R Core Team, 2021). R-code for all analyses steps as well as R-output can be found in the online supplemental material.

### 2.4.2. Statistical testing

First, we tested whether the D, rD, and noD groups differed in terms of their demographic or clinical characteristics, using analyses of variance (ANOVA) and  $\chi^2$ -tests. Where ANOVA assumptions were violated we used Kruskal-Wallis tests.

Second, we investigated differences between the three groups' PA

and NA distributions, by comparing their distributions of within-person means, standard deviations, and skewness values. We provide graphs to illustrate differences.

Third, to replicate the previous ESM studies of affective reactivity listed in Table 1, we adopted an analysis strategy that as closely as possible mirrored the analytical approach used in those studies. Accordingly, we used linear multilevel regression models to account for the hierarchical structure of the data, with repeated observations (level 1) clustered within individuals (level 2). Affective reactivity was modelled as the effect of momentary events on affect at the same measurement point. The four combinations of (positive and negative) event and (positive and negative) affect variables were analyzed in separate models. All estimated models included random intercepts and slopes and an unstructured covariance matrix for the random effects. Analyses were performed in a step-by-step manner. We first established whether there was a significant main effect of the event on the affective outcome. If this was the case, we further examined whether the effect differed between groups by including a group\*event cross-level interaction. All models included an autoregressive (lag-1) level-1 effect of the respective affective outcome (person-mean centered), to account for temporal dependency in the outcome.

Although not done in most previous studies, the event variables were included in a way that allowed for estimation of a pure within-person effects, rather than a combination of between- and within-person effects (Curran and Bauer, 2011; Enders and Tofighi, 2007). This was done by including person-mean centered event scores as a level-1 predictor and the person mean of event as a level-2 predictor. The former can be interpreted as the momentary effect of events on affect within persons—the effect of interest. The latter can be interpreted as the effect of between-person differences in event ratios on affect (we term this "event load" in the remainder of the paper). Most previous ESM studies (i.e., those included in Table 1 except Nelson et al., 2020; van der Stouwe et al., 2019) did not separate these effects, leaving it ambiguous whether their affective reactivity coefficients represented the within-person effects of momentary events or effects of between-person differences in event load (also see Cole et al., 2021). To obtain coefficients that reflect within-person effects, we intentionally deviated from previous studies on this aspect of the analysis. For group comparisons we included two dummy variables for the D and rD groups (with the noD group functioning as the reference group) as level-2 predictors, as well as their cross-level interactions with the event variable. To include comparisons between the D and rD groups, models were re-run with the D group as the reference group (the same was done with models in step 4). Analyses in this step were conducted using the R-packages 'lme4' (Bates et al., 2015) and 'lmerTest' (Kuznetsova et al., 2017). In lme4, missing data are handled using full information maximum likelihood estimation, assuming that missing observations are missing at random.<sup>2</sup>

Fourth, we examined whether the models investigating group differences in step 3 violated assumptions of multilevel regression. Specifically, we checked whether residual distributions were normal and whether residuals exhibited homoscedasticity (homogeneous variance over all predicted scores of the dependent variable). To this end, we created residual plots using the DHARMA package (Harting, 2021), which can be used to create residual plots for both linear and generalized linear (multilevel) models that can be interpreted analogous to residual plots for linear regression irrespective of the used link function. This is achieved in DHARMA by using randomized quantile residuals, which are continuous, even for discrete outcomes and have been shown to be approximately normally distributed under a correctly specified model

<sup>2</sup> Note that the analysis strategy in this paragraph does not mirror analyses in all prior ESM studies one to one, as prior ESM studies differ in certain analysis choices. For example, studies varied with regards to whether lagged affect was included in the model and whether two or three levels were modelled (including measurements within days as level 2).



**Table 2**  
Demographic and clinical characteristics of the sample by group.

|                                      | Non-depressed | Remitted     | Depressed   | Difference test                  |
|--------------------------------------|---------------|--------------|-------------|----------------------------------|
| N                                    | 96            | 189          | 61          |                                  |
| Age, mean (SD)                       | 50.1 (12.9)   | 47.8 (12.7)  | 52.0 (10.5) | $\chi^2(2) = 5.32, p = .070^a$   |
| Women, n (proportion)                | 63 (65.6 %)   | 131 (69.3 %) | 38 (62.3 %) | $\chi^2(2) = 1.15, p = .563^b$   |
| Years of education, mean (SD)        | 13.8 (3.0)    | 12.8 (3.0)   | 12.3 (3.1)  | $F(2, 343) = 5.18, p = .006^c$   |
| IDS-SR, mean (SD)                    | 6.6 (4.5)     | 14.1 (8.9)   | 30.7 (13.0) | $\chi^2(2) = 137.18, p < .001^a$ |
| BAI, mean (SD)                       | 2.7 (3.5)     | 7.4 (6.9)    | 14.6 (11.2) | $\chi^2(2) = 88.09, p < .001^a$  |
| Number of positive events, mean (SD) | 22.5 (16.7)   | 22.5 (14.7)  | 18.2 (13.7) | $\chi^2(2) = 4.43, p = .109^a$   |
| Number of negative events, mean (SD) | 6.2 (5.9)     | 9.1 (7.9)    | 9.1 (8.6)   | $\chi^2(2) = 12.76, p = .002^a$  |

Note. IDS-SR = Inventory of Depressive Symptomatology Self Report; BAI = Beck Anxiety Inventory.

<sup>a</sup> Kruskal-Wallis test.

<sup>b</sup>  $\chi^2$ -test of independence.

<sup>c</sup> ANOVA.

(Dunn and Smyth, 1996). More information on DHARMA is given in Harting (2021).

Fifth, in case of assumption violations, we proceeded to identify a better specified model, for which assumptions are met. To this end, we took several, increasingly complex steps, each time estimating the model and checking the model assumptions to see if they were met. The first step was to run the original model with a transformed outcome (NA or PA) variable. Specifically, we employed logarithmic transformation. If this proved ineffective to improve the residual distributions, we proceeded to the next step, which involved the use of generalized multilevel models. We started with models that are typically used to analyze continuous outcomes with non-normal distributions (Faraway, 2016): we ran generalized linear multilevel regression models using either a gamma or inverse Gaussian outcome distributions and a logarithmic link function in lme4 (Bates et al., 2015). If this did not result in normally and homogeneously distributed residuals, we proceeded to models that, instead of dealing with skewness by modeling a non-normal distribution, deal with a mass of values at zero. The models are referred to as two-part models for semi-continuous data (Farewell et al., 2017). The idea behind two-part models is to separately estimate whether the dependent variable is zero or not (part 1: binary) and if not zero, how much the variable differs from zero (part 2: continuous). Here, the dependent variable in question is considered semi-continuous in the sense that it arises from two processes: one explaining its presence versus absence (dichotomous) and one explaining the extent of its presence (continuous). In that regard, this model could be a good fit for an outcome that bottoms out at zero. Two-part models estimate two separate sets of parameters for the ‘binary’ and ‘continuous’ parts of the model and allow for estimation of a differently specified model for each part (e.g., one can include random effects for one part only). The binary part of the model can be viewed analogous to a logistic multilevel regression testing the odds of the dependent variable (Y) being zero, conditional on the independent variable(s). The continuous part of the model can be viewed analogous to a linear multilevel regression model estimating Y conditional on it being larger than zero and on the independent variable(s). Given the possibility of strong skewness of the outcome even when considering only the non-zero values, we specifically considered employment of a lognormal two-part models, in which non-zero outcome values were log-transformed prior to being included into the continuous part of the model. In order to control model complexity and aid model convergence, we introduced random effects one by one into the model, testing whether the random effect improved fit at each step via likelihood ratio tests. Two-part models were estimated using the ‘GLMMadaptive’ package (Rizopoulos, 2021). With the exception of random effects, parameter specifications matched the replication models from step 2 and were identical between the binary and continuous parts for all two-part models.

Analysis steps 3 and 5 included multiple tests. We applied the false discovery rate to correct *p*-values and thereby protect against false positive inflation in step 3 (Benjamini and Hochberg, 1995). We

refrained from a correction in step 5, since we predicted there to be no significant effect and a correction would have aided us in establishing absence of evidence. The cut-off for significance was set at  $\alpha = 0.05$ .

### 3. Results

#### 3.1. Sample characteristics

Demographic and clinical characteristics of our sample are given in Table 2. The D, rD, and noD groups did not differ on age and gender, but differed in the severity of depressive and anxiety symptoms. The groups also differed on the number of negative events, but not on the number of positive events.

#### 3.2. Are there floor/ceiling effects in NA or PA?

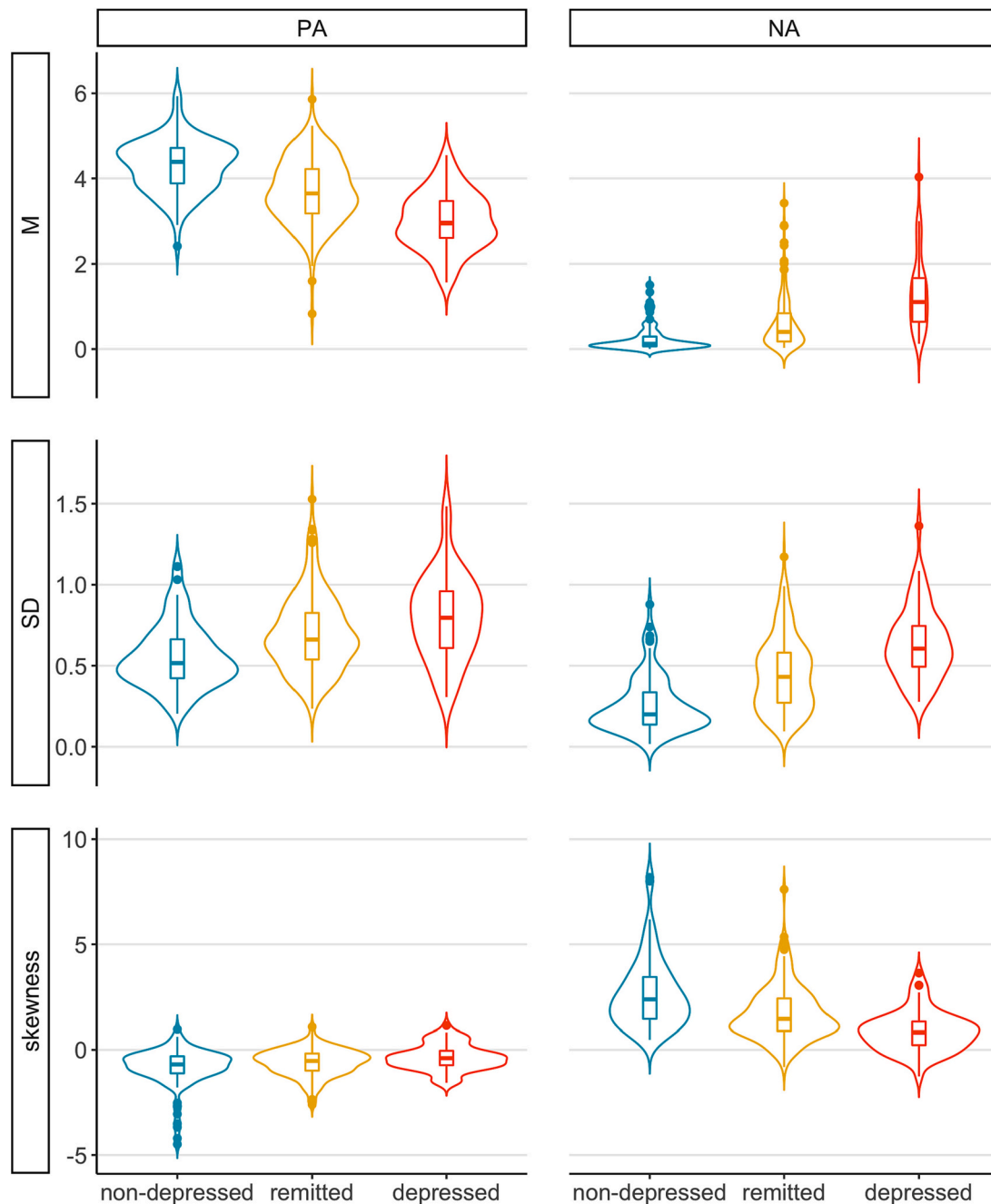
In Fig. 1 the distributions of within-person means, standard deviations, and skewness values of PA and NA are shown for the three groups. Group difference tests are included in S5 of the online supplemental material. While there are group differences in the distributions of both PA and NA, means were not positioned close to either bound of the scale for PA (Fig. 1). For NA however, there were a large number of values at the lower bound of the scale in the noD group. This was further illustrated by the proportion of zero values on NA across persons and measurements (noD: 51.7 %; rD: 31.1 %; D: 10.8 %).

#### 3.3. Replication analysis on group differences in affective reactivity

Before investigating group differences, we established whether there was reactivity in PA and NA to positive and negative events when using analytical methods that had previously been used. Results showed significant fixed effects of positive and negative-event load and positive and negative momentary events on both PA and NA (all  $p < .004$ ). Positive momentary events and event load were associated with increased PA and reduced NA, while negative momentary events and event load were associated with decreased PA and increased NA. In S6.2 of the online supplemental material more detailed outcomes of these models are provided.

Results from the naive replication models investigating group differences are shown in Table 3. For PA, positive momentary events had significantly stronger effects in the D and rD group when compared to the noD group.<sup>3</sup> There were no significant group differences in the effect of negative momentary events on PA. For NA, both positive and negative

<sup>3</sup> To check whether the increased momentary reactivity in the D group might be influenced by the lower rate of positive events in this group we re-ran the respective model including the interaction between positive event person means and person-mean centered positive events (see S11 in the online supplemental material). The reported difference in momentary reactivity persisted.



**Fig. 1.** Violin plots of within-person means (M), standard deviations (SD), and skewness of positive and negative affect (PA and NA, respectively) in the three groups. Note. Outside shapes give the mirrored density function and enclose box plots, where the central thick line indicates the median, the lower and upper end of the rectangle respectively indicate the 25th and 75th percentile, vertical lines indicate values that extend beyond these percentiles by maximally 1.5 \* the inter-quartile range, and points indicate values beyond this range of the vertical lines.

momentary events had a significantly stronger effect in the D group than in the noD group. The effect of positive and negative momentary events on NA was also significantly stronger in the rD group than in the noD group, while there was no significant difference between the D and rD groups. Results for the models with the extra comparison between the D and rD groups are provided in S6.4 of the online supplemental material.

### 3.4. Quantile residual distributions of replication models

Fig. 2 (first row) shows the residual plots for the naive replication models on negative events. Plots for the respective models on positive events were very similar to those in Fig. 2 and are included in S7 of the

online supplemental material. The plots show that the randomized quantile residuals from the NA replication models were not normally and homogeneously distributed. Residual plots for PA replication models showed approximately normally distributed residuals and homogeneous residual variances.

### 3.5. Analysis of affective reactivity accounting for affect distribution

Neither linear multilevel models of log-transformed NA nor generalized linear multilevel models with gamma or inverse-Gaussian outcome distributions had normally and homogeneously distributed residuals (Fig. 2). Model output for these models are included in S8 of

**Table 3**  
Replication model results for PA and NA reactivity to positive and negative events.

|                                 | Positive events                 |               |             | Negative events                 |               |             |
|---------------------------------|---------------------------------|---------------|-------------|---------------------------------|---------------|-------------|
|                                 | Parameter                       | Estimate (SE) | p-value     | Parameter                       | Estimate (SE) | p-value     |
| PA                              | Intercept                       | 4.30 (0.12)   | < .001      | Intercept                       | 4.52 (0.10)   | < .001      |
|                                 | Positive event (pmc)            | 0.24 (0.03)   | < .001      | Negative event (pmc)            | -0.50 (0.06)  | < .001      |
|                                 | PA <sub>t-1</sub> (pmc)         | 0.31 (0.01)   | < .001      | PA <sub>t-1</sub> (pmc)         | 0.33 (0.01)   | < .001      |
|                                 | Positive event (pm)             | 0.10 (0.27)   | .722        | Negative event (pm)             | -1.99 (0.77)  | .014        |
|                                 | rD group                        | -0.68 (0.15)  | < .001      | rD group                        | -0.54 (0.13)  | < .001      |
|                                 | D group                         | -1.57 (0.19)  | < .001      | D group                         | -1.52 (0.16)  | < .001      |
|                                 | rD group * positive event (pmc) | 0.13 (0.04)   | .003        | rD group * negative event (pmc) | -0.07 (0.07)  | .312        |
|                                 | D group * positive event (pmc)  | 0.19 (0.05)   | < .001      | D group * negative event (pmc)  | -0.15 (0.09)  | .111        |
|                                 | rD group * positive event (pm)  | 0.28 (0.35)   | .442        | rD group * negative event (pm)  | 0.40 (0.87)   | .660        |
|                                 | D group * positive event (pm)   | 1.00 (0.50)   | .059        | D group * negative event (pm)   | 2.39 (1.00)   | .023        |
|                                 | Model BIC                       | 34,007.36     |             | Model BIC                       | 33,567.28     |             |
|                                 | NA                              | Intercept     | 0.32 (0.10) | .002                            | Intercept     | 0.21 (0.09) |
| Positive event (pmc)            |                                 | -0.07 (0.02)  | .005        | Negative event (pmc)            | 0.28 (0.04)   | < .001      |
| NA <sub>t-1</sub> (pmc)         |                                 | 0.30 (0.01)   | < .001      | NA <sub>t-1</sub> (pmc)         | 0.30 (0.01)   | < .001      |
| Positive event (pm)             |                                 | -0.18 (0.15)  | .442        | Negative event (pm)             | 0.61 (0.64)   | .366        |
| rD group                        |                                 | 0.38 (0.12)   | .003        | rD group                        | 0.21 (0.11)   | .065        |
| D group                         |                                 | 1.18 (0.15)   | < .001      | D group                         | 1.07 (0.14)   | < .001      |
| rD group * positive event (pmc) |                                 | -0.10 (0.03)  | < .001      | rD group * negative event (pmc) | 0.14 (0.05)   | .013        |
| D group * positive event (pmc)  |                                 | -0.16 (0.04)  | < .001      | D group * negative event (pmc)  | 0.25 (0.07)   | < .001      |
| rD group * positive event (pm)  |                                 | -0.12 (0.28)  | .673        | rD group * negative event (pm)  | 0.70 (0.72)   | .366        |
| D group * positive event (pm)   |                                 | -0.75 (0.40)  | .074        | D group * negative event (pm)   | -1.03 (0.83)  | .241        |
| Model BIC                       |                                 | 21,263.67     |             | Model BIC                       | 20,030.44     |             |

Note. pmc = person-mean centered; pm = person mean; BIC = Bayesian Information Criterion.

the online supplemental material.

For the two-part models, the step-by-step introduction of random effects (all models are included in S9 of the online supplemental material) ultimately led to models including random intercepts and slopes in the continuous part of the model, as well as a random intercept in the binary part of the model, with all random-effect covariances being freely estimated. We refrained from further addition of random slopes in the binary part of the models, as it resulted in convergence problems that we were unable to resolve. Results for two-part models are given in Table 4. There were no significant group differences in the effect of momentary positive or negative events on NA in both the dichotomous and the continuous parts of the models, with one exception: positive momentary events had a stronger effect in the rD group than in the noD group in the continuous part of the model. Residual plots for the two-part models showed approximately normally distributed residuals and homogeneous residual variances (Fig. 2). Results for the models with the extra comparison between the D and rD groups are given in S9.5 of the online supplemental material.

While residual plots of PA replication models did not show clear assumption violations, we ran models on power-transformed PA to check if further improvement could be achieved (included in S10 of the online supplemental material). The resulting models also did not appear to violate assumptions, but model fit worsened considerably in comparison to models of non-transformed PA.

#### 4. Discussion

Findings from prior ESM studies investigating momentary affect indicated that depressed individuals react more strongly on NA and PA to both positive and negative stimuli than non-depressed individuals, particularly on NA. We hypothesized that extreme skewness due to floor effects in the NA-score distributions in non-depressed individuals could have influenced previously observed statistical results on affective reactivity, offering an alternative explanation for these previously observed group differences. We specifically hypothesized that NA floor effects in non-depressed individuals may influence affective reactivity estimates as they involve both restricted variability and non-normality of the NA distribution. This can lead to violations of the assumptions of commonly used statistical models and to the observation of spurious differences from groups with psychopathology, which do not exhibit a

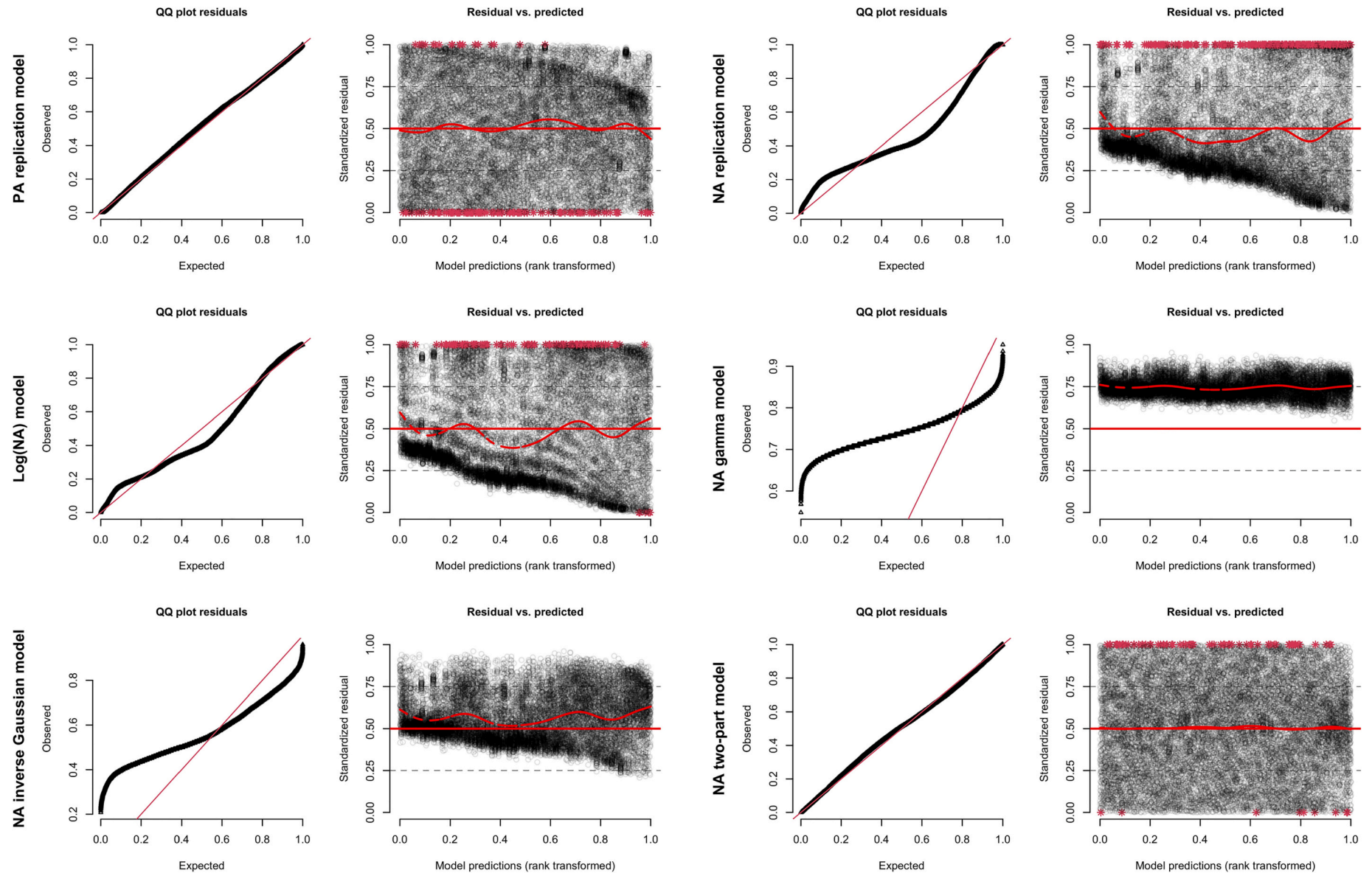
floor effect in their NA distributions. Our results support this alternative hypothesis.

We compared the effects of positive and negative events on PA and NA in three groups: depressed, remitted, and non-depressed individuals. We will first discuss our findings on NA reactivity and floor effects, then discuss our findings on PA as well as the remitted group, and finally reflect on limitations and future research.

Regarding NA, we found a very low mean in the non-depressed group and a strongly positively skewed distribution with a peak at the lowest possible score. This could suggest a floor effect due to a restriction in the range at the lower bound of the NA measure and/or a floor effect due to a realistic complete absence of NA in non-depressed individuals. Analyzing the data analog to previous ESM studies, without addressing the non-normality of NA, replicated previous findings of increased NA reactivity to positive and negative events in the depressed compared to the non-depressed group. Inspection of the residual distributions from this analysis indicated violations of model assumptions. In a better-specified model given the distributional characteristics of NA, no differences in NA reactivity between the depressed and non-depressed groups were found. These results show that previous affective reactivity findings were replicable, but that the observed group differences in NA reactivity did not hold once assumption violations were addressed. This indicates that the floor effect in NA-score distributions of non-depressed participants could explain at least part of the findings in the replication models. The influence of this floor effect was especially apparent in NA reactivity to positive events, as the low NA values in non-depressed individuals left little room for reductions in NA, distorting any comparison with a group that had room for reductions. Although the current findings are based on a single dataset and cannot be expected to directly generalize to previously conducted studies, it is plausible that previous studies on NA reactivity may have been similarly affected by non-normality and heteroscedasticity. Prior ESM studies on NA reactivity indeed reported similar NA distributions for their non-depressed groups in their sample descriptives (see Table S1 in the online supplemental material).

Although our results ultimately showed no group differences in NA reactivity, they showed the typical pronounced differences in the distributions of NA, with the depressed group reporting higher means and variances in NA than the non-depressed group (Telford et al., 2012). The idea that these basic distributional characteristics sufficiently describe





**Fig. 2.** Pairs of residual plots of models for negative events.

Note. QQ plots in the first and third column compare the distribution of observed model residuals against the distribution they are assumed to follow under the model. Divergences from the red diagonal line indicate divergences of the residuals from their expected distribution. Residual vs. predicted plots in the second and fourth column show the residuals against the value predicted by the model. The curved red line represents the estimated mean of residuals (via a smooth spline function). The horizontal red line at 0.5 indicated the expected mean of residuals. Values plotted in red are simulation outliers. To fulfill model assumptions residuals should be approximately uniformly distributed around a mean of 0.5 and with homogeneous variance across levels of predicted values. See [Harting \(2021\)](#) for more information on residual plots created with DHARMA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 4**  
Two-part model results for NA reactivity to positive and negative events.

| Parameter                       | Positive events  |         |                 |         | Parameter                       | Negative events  |         |                 |         |
|---------------------------------|------------------|---------|-----------------|---------|---------------------------------|------------------|---------|-----------------|---------|
|                                 | Dichotomous part |         | Continuous part |         |                                 | Dichotomous part |         | Continuous part |         |
|                                 | Estimate (SE)    | p-value | Estimate (SE)   | p-value |                                 | Estimate (SE)    | p-value | Estimate (SE)   | p-value |
| Intercept                       | −0.06 (0.46)     | .898    | −1.04 (0.11)    | < .001  | Intercept                       | 0.78 (0.39)      | .047    | −1.31 (0.09)    | < .001  |
| Positive event (pmc)            | 0.79 (0.10)      | < .001  | −0.08 (0.04)    | 0.061   | Negative event (pmc)            | −1.57 (0.14)     | < .001  | 0.43 (0.05)     | < .001  |
| NA <sub>t-1</sub> (pmc)         | −1.36 (0.07)     | < .001  | 0.36 (0.02)     | < .001  | NA <sub>t-1</sub> (pmc)         | −1.38 (0.07)     | < .001  | 0.37 (0.02)     | < .001  |
| Positive event (pm)             | 0.93 (1.08)      | .385    | −0.44 (0.26)    | .083    | Negative event (pm)             | −5.03 (2.97)     | .090    | 0.83 (0.68)     | .226    |
| rD group                        | −1.77 (0.59)     | .003    | 0.43 (0.14)     | .001    | rD group                        | −1.77 (0.49)     | < .001  | 0.49 (0.12)     | < .001  |
| D group                         | −4.91 (0.78)     | < .001  | 1.15 (0.17)     | < .001  | D group                         | −5.29 (0.65)     | < .001  | 1.36 (0.14)     | < .001  |
| rD group * positive event (pmc) | −0.11 (0.12)     | .377    | −0.10 (0.05)    | .035    | rD group * negative event (pmc) | 0.02 (0.18)      | .892    | 0.02 (0.06)     | .711    |
| D group * positive event (pmc)  | −0.14 (0.20)     | .468    | −0.10 (0.06)    | .090    | D group * negative event (pmc)  | 0.42 (0.30)      | .169    | 0.02 (0.07)     | .801    |
| rD group * positive event (pm)  | −0.32 (1.40)     | .819    | 0.13 (0.32)     | .682    | rD group * negative event (pm)  | 0.42 (3.40)      | .901    | −0.20 (0.78)    | .801    |
| D group * positive event (pm)   | 1.29 (2.03)      | .527    | −0.13 (0.44)    | .772    | D group * negative event (pm)   | 6.43 (3.91)      | .100    | −1.58 (0.87)    | .071    |
| Model BIC                       | 33,952.08        |         |                 |         | Model BIC                       | 33,314.22        |         |                 |         |

Note. pmc = person-mean centered; pm = person mean; BIC = Bayesian Information Criterion. Note that the dichotomous part estimates the chance of the outcome (NA) being zero, so that, for example, positive parameter estimates indicate that increases in the predictor (e.g., positive event) predict increases in the number of zeros.

affect fluctuations in depression has been introduced by Dejonckheere et al. (2019). Dejonckheere and colleagues showed that various parameters describing affect dynamics (such as affective instability, inertia, and differentiation) are empirically conflated with each other and do not uniquely contribute to explaining depression scores over and above the differences in mean and variance of affect scores. Our results extend this point to NA reactivity, which also emerged as unrelated to depression (here specifically depression diagnosis), if one accounts for group differences in NA distributions. A previous analysis of the same dataset by Schoevers et al. (2020) indicated that differences in the distributions indeed pertain to both mean and variance. They showed that differences in affective instability, a metric closely related to the variance, persisted when accounting for mean differences. The overall lesson seems to be that, when we study affective reactivity and other affect dynamics, it is important to take into account the affect distributions and to select a well specified statistical modeling approach.

In our results regarding PA, the distributions (i.e., means, standard deviations, and skewness) did not indicate any floor/ceiling effects and replication models did not violate statistical assumptions. These models indicated that PA reactivity to positive events was increased in the depressed group compared to the non-depressed group, but showed no difference for negative events. The increased reactivity to positive events in depressed individuals is in line with two prior ESM studies that found a similar ‘mood-brightening’ effect in PA (Khazanov et al., 2019; Peeters et al., 2003). However, there are also four previous studies that found no evidence for this effect (Bylsma et al., 2011; Heininga et al., 2019; Nelson et al., 2020; Thompson et al., 2012; also see Table 1) and two studies that found no group difference in PA reactivity to positive activities (Heininga et al., 2019; Nelson et al., 2020). Overall, findings with regard to the PA-mood-brightening effect have so far been inconsistent. In contrast, all ESM studies, including the current one, have found no group differences in PA reactivity to negative events (Bylsma et al., 2011; Nelson et al., 2020; Thompson et al., 2012; van Winkel et al., 2015), except for Peeters et al. (2003), who found reduced PA reactivity to negative events in depressed compared non-depressed individuals. Taken together, our results on PA and NA reactivity differ considerably from findings in prior ESM studies. These findings raise uncertainty about the question whether depressed individuals are indeed characterized by divergent affective reactivity.

Our study provides novel insights into affective reactivity in individuals remitted from depression. Individuals in this group did not meet the criteria of a depression diagnosis anymore, but on average showed more depressive symptoms than individuals in the non-depressed group. Remitted individuals showed no differences in affective reactivity from the depressed group. Compared to the non-

depressed group they did not differ in reactivity to negative events, but showed increased PA and NA reactivity to positive events. The difference in NA reactivity should be interpreted very cautiously. Even using two-part models, it remains problematic to investigate reductions in NA (i.e., in response to positive events) where there is restricted room for such reductions, as was the case in our non-depressed group. The finding of increased PA reactivity to positive events among remitted individuals does not suffer from this limitation. Instead, it aligns with our finding of a PA-mood-brightening effect in depressed individuals. To our knowledge, the only previous study that included remitted individuals also found no evidence for differences in PA reactivity to three types of negative stimuli, when compared to a non-depressed group (van Winkel et al., 2015). They did however report increased NA-reactivity to one type of stimulus (i.e., social stress). However, these previous results should be interpreted carefully, taking into account the limited number of remitted participants in their study ( $N = 11$ ) and the potential influence of a non-normal NA distribution on the observed effects.

The following limitations of our study should be considered. First, although our findings showcase how a floor effect in the NA distribution of healthy individuals can influence findings of NA reactivity and while we observe in the descriptive statistics of many previous studies that floor effects were likely present, we do not know to what extent their results were affected by floor effects. Only a re-analysis of the original data from these studies could clarify this. Strictly speaking, we also did not exactly reproduce all prior ESM studies. There are many aspects in which these studies varied from our study, such as the investigated stimuli, the items that PA- and NA-scales are composed of, the ESM designs, and the analyses. It is unclear to what extent any of these variations would be expected to contribute to observed differences in the results. Second, we would like to stress that the two-part models we employed in our analysis are not a general solution for analyzing data that contain floor effects. As we point out above, under these models there is still restricted room for NA reductions with a floor effect. Restricted variability thus remains an issue. Two-part models can effectively account for a large number (peak) of zero values in the outcome variable, which are assumed to be ‘true’ zeros. Yet, a floor effect could also be caused by problems with the measurement instrument, where the instrument does not cover the full range of the phenomenon in question, leading to zeros that are not true zeros but proxies for low-level expressions of the phenomenon that fall below the lower bound of the measurement range. If this is the case, this is a measurement issue that affects the interpretability of the results irrespective of the used analytical method, including two-part models. In our study, zeros on NA resulted from participants choosing the minimum answer with the anchor ‘not at all’ on multiple NA items. In effect, this could

mean that participants actually reported zero NA. However, it is also conceivable that zeros were reported because ESM questions i) did not differentiate enough between lower levels of NA expression (e.g., too much difference between the lowest and second to lowest response) and/or ii) did not cover lower-level expressions of NA that lie on the non-pathological side of the affective spectrum. Lastly, it is important to stress that our results only pertain to the question whether divergent affective reactivity is a characteristic of depressed or remitted individuals versus non-depressed individuals. It does not pertain to the role of variations in affective reactivity as a risk factor for depression onset (Wichers et al., 2007a, 2007b; Wichers et al., 2009), as a predictor of depression course (Peeters et al., 2010; Wichers et al., 2010), or as a characteristic of individuals with non-clinical expressions of depression in population samples (Booij et al., 2018; Santee and Starr, 2021; van Roekel et al., 2016). Neither do our results pertain to divergences on other metrics of affect dynamics as a characteristic of depressed individuals (for an overview, see Kuppens and Verduyn, 2017).

Based on the finding that floor effects have an impact on estimates of NA reactivity, we recommend that researchers pay extra attention to their study design and statistical approach when investigating affective reactivity in ESM studies. We recommend that, both in potential re-analysis of previous studies and in future studies on affective reactivity, the influence of floor effects and other distributional deviations in the used outcome variables is evaluated and addressed. First, one can consider during study design whether a given ESM measure is adequate and useful for a given population under investigation, especially when measuring expressions of psychopathology in healthy participants (Terluin et al., 2016). Given that about half of all NA measurements were scored zero in our non-depressed group, one might reasonably question whether NA (as we measured it) optimally operationalizes affect variations in this group. Here, it at least appears to be limited for capturing improvements in affect (i.e., reductions in NA). Future ESM studies might try to better capture low-level expressions of NA, or focus on PA, which does not seem to be limited in the same way. Second, once data are collected, outcome distributions as well as the model assumptions should be tested and reported in the manuscript. Checking assumptions is absolutely essential for judging the appropriateness of a model and only through reporting these checks and subsequent actions can researchers do justice to the principle of transparency in science (Ernst and Albers, 2017). If necessary, transformations of skewed outcome variables or alternative, better-specified models, for example two-part models, can be applied. Note that it cannot be assumed that applying any of these approaches will solve assumption violations the first time around, so this should be checked again before making any inferences based on the model. Also note that different outcome variables in the same study might require different modeling approaches due to differences in their distributions (e.g., NA and PA in the current sample). This can make it harder to directly compare the results across outcome variables, which may be inconvenient when such direct comparisons are important from a researcher's theoretical or clinical perspective. Additionally, given that results on affective reactivity, as is the case for many ESM-based results, are usually interpreted on the within-subject level, researchers should ideally use the recommended analytical approaches for separating within- and between-person variation. This entails person-mean centering the independent variable in the level-1 model and including the independent variable's person-mean in the level-2 model (e.g., Bolger and Laurenceau, 2013; Cole et al., 2021; Curran and Bauer, 2011; Kreft et al., 1995). Only then, the level-1 coefficient of the independent variable (e.g., event) can be interpreted as a purely within-person effect.

Given the overall inconsistency in previous and present findings, it remains uncertain if depressed and non-depressed individuals do or do not differ in their affective reactivity. Although further research will be necessary to unravel the inconsistencies, we would like to suggest some possible explanations. We already mentioned possible issues with the way in which affect is measured. Another explanation might be that the

operationalization of stimuli (e.g., events) as unidimensional (or even dichotomous) variables involves a high degree of simplification, making it hard to detect more specific relevant signals from among the diverse range of environments that participants encounter in their daily lives. Additionally, these stimulus variables are not direct representations of stimuli in the environment, but represent participants' appraisal of such stimuli. To the degree that individuals appraise their environments subjectively (e.g., having different thresholds or idiosyncratic vulnerabilities), the stimulus variables represent different things for different participants, making them less suited for direct between-person comparisons. A potential path for ESM research to investigate reactions to more unambiguous stimuli would be to take a quasi-experimental approach, in which the reaction to the same naturally occurring event is tracked (Dejonckheere et al., 2021; Kalokerinos et al., 2019), or an experimental approach, in which responses to a stress task are repeatedly measured in participants' daily lives. Additionally, one could focus on the diverse stimuli and their subjective interpretation by capturing them in a more qualitative manner, for instance by asking participants to describe the stimulus by typing in a text window. The question "what (subjective) stimuli do depressed and non-depressed individuals react strongly to?" could offer an alternative perspective on affective reactivity.

Unraveling the inconsistencies in ESM findings on affective reactivity and addressing the abovementioned concerns regarding its operationalization will be necessary if we want to draw conclusions about how depressed individuals emotionally interact with their daily-life environments. Such conclusions are further complicated by the fact that ESM findings of increased reactivity in depressed individuals are at odds with findings from lab studies, which show the opposite pattern, namely that depressed individuals are less responsive to stimuli than healthy controls (a finding that has been termed *emotional context insensitivity*; Bylsma et al., 2008; Rottenberg, 2017). Clearly, these opposing patterns of findings also have different clinical implications. We therefore recommend that ESM research remains cautious with interpreting findings clinically and prioritizes resolving the abovementioned issues.

In summary, the present study showed that floor effects may lead to an overestimation of the increased NA reactivity in depressed versus non-depressed individuals that was repeatedly found in ESM studies. Second, and contrary to most previous studies, we found an increased PA-mood-brightening effect in depressed and remitted, compared to non-depressed participants. In light of these findings, it appears that ESM research currently does not have a definitive answer to the question whether depressed individuals are indeed characterized by divergent affective reactivity in their daily lives. Our findings open new opportunities for future research and remind us that choices in research design and statistical methods can affect research findings and the conclusions we draw.

#### **CRedit authorship contribution statement**

All co-authors have substantially contributed to the manuscript. All have read and approved the final version of the manuscript.

#### **Declaration of competing interest**

We declare no conflicts of interest.

#### **Acknowledgements**

This work was supported by the charitable foundation Stichting tot Steun VCVGZ (awarded to HR, grant 239) and the foundation Researchfonds Stichting Postacademische Psy-opleidingen (awarded to MNS, grant PPO-MS-MH). The infrastructure for the NESDA study ([www.nesda.nl](http://www.nesda.nl)) has been funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and by participating universities



and mental health care organizations (Amsterdam University Medical Centers (location VUmc), GGZ inGeest, Leiden University Medical Center, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Dimence Groep, Rob Giel Onderzoekcentrum).

## Appendix A. Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.jad.2023.04.118>.

## References

- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. American Psychiatric Publishing.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* 56, 893–897.
- 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, 289–300.
- Bolger, N., Laurenceau, J.-P., 2013. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. Guilford Press, New York, NY.
- Booij, S.H., Snippe, E., Jeronimus, B.F., Wichers, M., Wigman, J.T.W., 2018. Affective reactivity to daily life stress: relationship to positive psychotic and depressive symptoms in a general population sample. *J. Affect. Disord.* 225, 474–481.
- Bylsma, L.M., Morris, B.H., Rottenberg, J., 2008. A meta-analysis of emotional reactivity in major depressive disorder. *Clin. Psychol. Rev.* 28, 676–691.
- Bylsma, L.M., Taylor-Clift, A., Rottenberg, J., 2011. Emotional reactivity to daily events in major and minor depression. *J. Abnorm. Psychol.* 120, 155–167.
- Cole, D.A., Gabruk, M.E., Nestor, B.A., Liu, Q., Lovette, A.J., Zhang, Y., Mick, C.R., 2021. Between- versus within-person emotional and cognitive reactivity in relation to depressive symptoms. *J. Affect. Disord.* 295, 479–487.
- Curran, P.J., Bauer, D.J., 2011. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* 62, 583–619.
- Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., Tuerlinckx, F., 2019. Complex affect dynamics add limited information to the prediction of psychological well-being. *Nat. Hum. Behav.* 3, 478–491.
- Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., Kalokerinos, E.K., 2021. The relation between positive and negative affect becomes more negative in response to personally relevant events. *Emotion* 21, 326–336.
- Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. *J. Comput. Graph. Stat.* 5, 236–244.
- Enders, C.K., Tofghi, D., 2007. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* 12, 121–138.
- Ernst, A.F., Albers, C.J., 2017. Regression assumptions in clinical psychology research practice—a systematic review of common misconceptions. *PeerJ* 5, e3323.
- Faraway, J.J., 2016. *Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 2nd ed. CRC Press.
- Farewell, V.T., Long, D.L., Tom, B.D.M., Yiu, S., Su, L., 2017. Two-part and related regression models for longitudinal data. *Ann. Rev. Stat. Appl.* 4, 283–315.
- Fries, J.F., Lingala, B., Siemons, L., Glas, C.A.W., Cella, D., Hussain, Y.N., Bruce, B., Krishnan, E., 2014. Extending the floor and the ceiling for assessment of physical function. *Arthritis Rheumatol.* 66, 1378–1387.
- Gelman, A., Hill, J., 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.
- Harting, F., 2021. *DHARMA: Residual Diagnostics for Hierarchical (Multi-level/Mixed) Regression Models*, 0.4.3 ed.
- Heininga, V.E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., Roy, J. V., Kuppens, P., 2019. The dynamical signature of anhedonia in major depressive disorder: positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry* 19, 59.
- Hoffmann, J.P., 2021. *Linear Regression Models: Applications in R*. CRC Press, Boca Raton.
- Houben, M., van den Noortgate, W., Kuppens, P., 2015. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. *Psychol. Bull.* 141, 901–930.
- Jacobson, N.S., Martell, C.R., Dimidjian, S., 2001. Behavioral activation treatment for depression: returning to contextual roots. *Clin. Psychol. Sci. Pract.* 8, 255–270.
- Kalokerinos, E.K., Erbas, Y., Ceulemans, E., Kuppens, P., 2019. Differentiate to regulate: low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychol. Sci.* 30, 863–879.
- Khazanov, G.K., Ruscio, A.M., Swendsen, J., 2019. The “Brightening” effect: reactions to positive events in the daily lives of individuals with major depressive disorder and generalized anxiety disorder. *Behav. Ther.* 50, 270–284.
- Kreft, I.G.G., de Leeuw, J., Aiken, L.S., 1995. The effect of different forms of centering in hierarchical linear models. *Multivariate Behav. Res.* 30, 1–21.
- Kuppens, P., Verduyn, P., 2017. Emotion dynamics. *Current Opinion in Psychology* 17, 22–26.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. *Applied Linear Statistical Models*, 5 ed. McGraw-Hill Irwin, New York.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26.
- Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., Tuerlinckx, F., 2018. Sideline the mean: the relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychol. Methods* 23, 690–707.
- Montgomery, D.C., Peck, E.A., Vining, G.G., 2021. *Introduction to Linear Regression Analysis*, 6 ed. John Wiley & Sons, New York.
- Muntings, A.D., Van Der Feltz-Cornelis, C.M., Van Marwijk, H.W., Spinhoven, P., Penninx, B.W., Van Balkom, A.J., 2011. Is the beck anxiety inventory a good tool to assess the severity of anxiety? A primary care study in the Netherlands study of depression and anxiety (NESDA). *BMC Fam. Pract.* 12, 66.
- Myin-Germeys, I., Peeters, F., Havermans, R., Nicolson, N.A., de Vries, M.W., Delespaul, P., Os, J.V., 2003. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr. Scand.* 107, 124–131.
- Nelson, J., Klumparendt, A., Doebler, P., Ehring, T., 2020. Everyday emotional dynamics in major depression. *Emotion* 20, 179–191.
- Panaite, V., Whittington, A., Cowden Hindash, A., 2018. The role of appraisal in dysphoric affect reactivity to positive laboratory films and daily life events in depression. *Cognit. Emot.* 32, 1362–1373.
- Panaite, V., Koval, P., Dejonckheere, E., Kuppens, P., 2019. Emotion regulation and mood brightening in daily life vary with depressive symptom levels. *Cognit. Emot.* 33, 1291–1301.
- Peeters, F., Nicolson, N.A., Berkhof, J., Delespaul, P., deVries, M., 2003. Effects of daily events on mood states in major depressive disorder. *J. Abnorm. Psychol.* 112, 203–211.
- Peeters, F., Berkhof, J., Rottenberg, J., Nicolson, N.A., 2010. Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. *Behav. Res. Ther.* 48, 754–760.
- Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.J., Van Der Meer, K., Verhaak, P., Wensing, M., De Graaf, R., Hoogendijk, W.J., Ormel, J., Van Dyck, R., 2008. The Netherlands study of depression and anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* 17, 121–140.
- Penninx, B.W.J.H., Eikelenboom, M., Giltay, E.J., Van Hemert, A.M., Riese, H., Schoevers, R.A., Beekman, A.T.F., 2021. Cohort profile of the longitudinal Netherlands study of depression and anxiety (NESDA) on etiology, course and consequences of depressive and anxiety disorders. *J. Affect. Disord.* 287, 69–77.
- R Core Team, 2021. *R: A Language and Environment for Statistical Computing*, 4.0.3 ed. R Foundation for Statistical Computing, Vienna, Austria.
- Rizopoulos, D., 2021. *GLMMadaptive: Generalized Linear Mixed Models Using Adaptive Gaussian Quadrature*, 0.8–0 ed.
- Rottenberg, J., 2017. Emotions in depression: what do we really know? *Annu. Rev. Clin. Psychol.* 13, 241–263.
- Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The inventory of depressive symptomatology (IDS): psychometric properties. *Psychol. Med.* 26, 477–486.
- Santee, A.C., Starr, L.R., 2021. Examining linear and nonlinear associations between negative emotional reactivity to daily events and depression among adolescents. *Clin. Psychol. Sci.* 10, 675–689.
- Schoevers, R.A., van Borkulo, C.D., Lamers, F., Servaas, M.N., Bastiaansen, J.A., Beekman, A.T.F., Van Hemert, A.M., Smit, J.H., Penninx, B.W.J.H., Riese, H., 2020. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. *Psychol. Med.* 1–10.
- Šimkovic, M., Träuble, B., 2019. Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLoS One* 14, e0220889.
- Sytema, S., van der Krieke, L., 2013. Routine outcome monitoring: a tool to improve the quality of mental health care. In: Thornicroft, G., Ruggeri, M., Goldberg, D. (Eds.), *Improving Mental Health Care: The Global Challenge*. John Wiley & Sons Ltd, pp. 246–263.
- Telford, C., McCarthy-Jones, S., Corcoran, R., Rowse, G., 2012. Experience sampling methodology studies of depression: the state of the art. *Psychol. Med.* 42, 1119–1129.
- Terluin, B., de Boer, M.R., de Vet, H.C.W., 2016. Differences in connection strength between mental symptoms might be explained by differences in variance: reanalysis of network data did not confirm staging. *PLoS One* 11, e0155205.
- Thompson, R.J., Mata, J., Jaeggi, S.M., Buschkuhl, M., Jonides, J., Gotlib, I.H., 2012. The everyday emotional experience of adults with major depressive disorder: examining emotional instability, inertia, and reactivity. *J. Abnorm. Psychol.* 121, 819–829.
- Thompson, R.J., Bailen, N.H., English, T., 2021. Everyday emotional experiences in current and remitted major depressive disorder: an experience-sampling study. *Clin. Psychol. Sci.* 9, 866–878.
- van der Stouwe, E.C.D., Groenewold, N.A., Bos, E.H., de Jonge, P., Wichers, M., Booij, S. H., 2019. How to assess negative affective reactivity to daily life stress in depressed and nondepressed individuals? *Psychiatry Res.* 279, 259–266.
- van Roekel, E., Bennis, E.C., Bastiaansen, J.A., Verhagen, M., Ormel, J., Engels, R.C.M.E., Oldehinkel, A.J., 2016. Depressive symptoms and the experience of pleasure in daily life: an exploration of associations in early and late adolescence. *J. Abnorm. Child Psychol.* 44, 999–1009.
- van Winkel, M., Nicolson, N.A., Wichers, M., Viechtbauer, W., Myin-Germeys, I., Peeters, F., 2015. Daily life stress reactivity in remitted versus non-remitted depressed individuals. *Eur. Psychiatry* 30, 441–447.

- Ward, M.M., Guthrie, L.C., Alba, M., 2014. Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *J. Clin. Epidemiol.* 67, 689–696.
- Watson, D., Tellegen, A., 1985. Toward a consensual structure of mood. *Psychol. Bull.* 98, 219–235.
- Wichers, M., Germeys, I.M., Jacobs, N., Peeters, F., Kenis, G., Derom, C., Vlietinck, R., Delespaul, P., Os, J.V., 2007. Evidence that moment-to-moment variation in positive emotions buffer genetic risk for depression: a momentary assessment twin study. *Acta Psychiatr. Scand.* 115, 451–457.
- Wichers, M., Germeys, I.M., Jacobs, N., Peeters, F., Kenis, G., Derom, C., Vlietinck, R., Delespaul, P., Os, J.V., 2007. Genetic risk of depression and stress-induced negative affect in daily life. *Br. J. Psychiatry* 191, 218–223.
- Wichers, M., Geschwind, N., Jacobs, N., Kenis, G., Peeters, F., Derom, C., Thiery, E., Delespaul, P., Os, J.V., 2009. Transition from stress sensitivity to a depressive state: longitudinal twin study. *Br. J. Psychiatry* 195, 498–503.
- Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C.J.P., Derom, C., Thiery, E., Delespaul, P.H., Os, J.V., 2010. Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: a momentary assessment study. *J. Affect. Disord.* 124, 191–195.
- Wittchen, H.-U., 1994. Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): a critical review. *J. Psychiatr. Res.* 28, 57–84.