Hindawi
Depression Research and Treatment
Volume 2019, Article ID 3481624, 9 pages
https://doi.org/10.1155/2019/3481624

Research Article

Heterogeneity Matters: Predicting Self-Esteem in Online Interventions Based on Ecological Momentary Assessment Data

Vincent Bremer[1], **Burkhardt Funk**[1] and **Heleen Riper**[2,3]

[1]*Institute of Information Systems, Leuphana University, Lueneburg, Germany*
[2]*Department of Clinical, Neuro- & Developmental Psychology, Vrije University, Amsterdam, Netherlands*
[3]*Amsterdam Department of Health Sciences, Vrije University, Amsterdam, Netherlands*

Correspondence should be addressed to Vincent Bremer; vincent.bremer@leuphana.de

Received 10 April 2018; Revised 24 November 2018; Accepted 25 December 2018; Published 13 January 2019

Academic Editor: Axel Steiger

Self-esteem is a crucial factor for an individual's well-being and mental health. Low self-esteem is associated with depression and anxiety. Data about self-esteem is oftentimes collected in Internet-based interventions through Ecological Momentary Assessments and is usually provided on an ordinal scale. We applied models for ordinal outcomes in order to predict the self-esteem of 130 patients based on diary data of an online depression treatment and thereby illustrated a path of how to analyze EMA data in Internet-based interventions. Specifically, we analyzed the relationship between mood, worries, sleep, enjoyed activities, social contact, and the self-esteem of patients. We explored several ordinal models with varying degrees of heterogeneity and estimated them using Bayesian statistics. Thereby, we demonstrated how accounting for patient-heterogeneity influences the prediction performance of self-esteem. Our results show that models that allow for more heterogeneity performed better regarding various performance measures. We also found that higher mood levels and enjoyed activities are associated with higher self-esteem. Sleep, social contact, and worries were significant predictors for only some individuals. Patient-individual parameters enable us to better understand the relationships between the variables on a patient-individual level. The analysis of relationships between self-esteem and other psychological factors on an individual level can therefore lead to valuable information for therapists and practitioners.

1. Introduction

Access to mental care is limited; by providing further access, Internet-based interventions can close the gap between treatment and demand [1–3]. At the same time, online-based interventions may lead to comparable outcomes compared to face-to-face treatment [1, 4]. In Internet-based interventions, data about various psychological factors, for example, the self-esteem level of individuals, is often collected. Self-esteem is closely related to psychological well-being and satisfaction with life [5]. Low levels of self-esteem are associated with serious mental problems such as depression, anxiety [6], or eating disorders [7]. Trzesniewski et al. [8] found that low self-esteem can lead to "negative real-world consequences" such as mental and physical health problems, misconduct, and worse economic outlooks. In the literature, however, there is a debate if low mood levels affect self-esteem or vice versa. Two models exist for each assumption. The vulnerability model assumes that self-esteem is a risk for depression whereas the scar model interprets self-esteem rather as an outcome or aftermath of depression [9]. One study, for example, found that low self-esteem can predict depression decades later [10]. Steiger et al. [11] found that the vulnerability and the scar model are valid over decades with weaker effects for the scar model. A reoccurring finding is that low levels of self-esteem are associated with serious mental illnesses which in turn are known to be associated with decreased quality of life and tremendous health care costs, as well as increased costs for individuals and governments [5–7, 12–14]. Thus, we aimed at predicting the self-esteem level of individuals in this study and analyze its relationships with a variety of psychological factors.

Data about self-esteem and other psychological factors such as mood levels or social interactions are often assessed by Ecological Momentary Assessments (EMA). These EMA methods collect data regarding behavior, symptoms, and

cognition close in time to the users' experience and in their natural environment [15, 16]. Diaries, which are used for the analysis in this paper, are one example of EMA methods that are often utilized [15].

Due to multiple measures per individual, this data has a nested structure [17, 18]. As is common in the social sciences [19], self-reports of diary data can be ranked on an ordinal scale. Individuals are often prompted to rank their mood level, for instance, by providing a score between one and ten for a specific question such as "*How is your mood right now?*" Data with this structure needs to be analyzed by utilizing appropriate statistical models that can account for the ordinality in the measurements, for example, ordinal logit models or generalized linear models. In research studies, however, this is often not the case [20–22]. Jakobsson [20] and LaValley and Felson [21] analyzed a multitude of journal articles; eventhough ordinal scales were often used, they came to the conclusion that frequently there were no appropriate data representation techniques or data analysis methods present. They found that solely 49% (La Valley et al.: 39.4%) of the analyzed articles had proper data presentation and 57% (La Valley et al.: 63.4%) had appropriate data analysis. This is alarming since an improper handling can lead to bias and incorrect interpretation of statistical effects [23].

Each patient behaves differently, has different experiences, and can be affected by psychological factors in various ways. Repeated measurements provided by patients can therefore not be considered to be independent [24]. Considering the differences among patients by implementing patient-individual parameters might lead to a better model fit (representation of the pattern in the data) and an increased prediction performance (ability to predict unobserved values of the dependent variable). By revealing these patient-individual parameters, individual effects for the independent variables (psychological factors) can be obtained for each patient, which in turn can result in individualized decision support systems and subsequently individualized recommendations in a clinical context.

In this study, we thus combined ordinal models appropriate for the analysis of diary data, namely, the ordinal logit model [25, 26] and the less frequently utilized stereotype logit model [26, 27], and proposed to extend the models by including patient-specific parameters in order to account for heterogeneity among the participants. General mixed models are often applied when analyzing data that includes repeated measurements [24]. Hedeker [23], for example, discussed mixed effects logistic regression models for ordinal data and illustrated a possible hierarchical structure in which the effect each patient has on the outcome value is considered. In contrast to this study, our approach considered different influences of the psychological factors on the individuals which led to individual slopes. These patient-specific coefficients can potentially result in more information on how the analyzed psychological factors are related to the self-esteem of the patients on an individual level and can therefore lead to a knowledge gain for researchers and practitioners. We applied the models to self-reported diary data from an Internet-based depression treatment [28] in order to predict the self-esteem of individuals. At the same time, we revealed the relationship between a variety of psychological/psychosocial factors (mood, worries, sleep, enjoyed activities, and social contact) and the self-esteem level of patients. Thus, this study contributes to existing research by gaining insight into the patients' behavior and how their self-esteem is related to a variety of factors on an individual level and thus by highlighting the importance of individuality in this context.

2. Materials and Methods

2.1. Data. The data we utilized for our approach is acquired from an EU funded two-arm randomized control trial that compared bCBT (blended cognitive behavior therapy, experiment group) and face-to-face treatment (control group) [28]. Participants were 18 years or older, met criteria for a major depressive disorder, were not of high suicidal risk, were not currently being treated for depression, and had access to an Internet connection. The utilized data was based on diary data that has been assessed in the study through an EMA mobile-application between February 2015 and January 2017. The diary questions were sent via email or text message depending on the therapists' choice. The mood level of the participants was collected every day at a random time between 10 a.m. and 8 p.m. All other factors were collected on specific days; the first and last seven days of the intervention and one random day each week in the intervention period. All factors could be ranked on a scale from one to ten. We only utilized days on which all factors were assessed, which resulted in the analysis of 130 patients and their 2326 observations including all psychological factors that will be introduced in the following.

Self-Esteem — The dependent variable in our analysis was the self-esteem of the patients. It was assessed through the question "How do you feel about yourself right now?" This question is closely related to an item of the state self-esteem scale [29] and can represent a person's self-image [30]. The same question has also been utilized in another study that measured self-esteem for individuals and has shown to be correlated with the Rosenberg self-esteem scale [31–33]. In this study, we defined this question as the self-esteem level.

Mood — Mood is an important factor for an individual's well-being, physical health, and behavioral patterns [34, 35]. We analyzed the relationship between these factors and hypothesized that the mood level is positively related to self-esteem. This predictor was assessed by the question "How is your mood right now?"

Worry — Worries are connected to anxiety disorders [36] and depression [37]. Since the act of worrying can potentially create feelings and thoughts that impact self-respect or cause individuals to underestimate themselves, it could be linked to self-esteem. We hypothesized that this factor is negatively related to the self-esteem of the patients. Worries were assessed by asking the patients "How much do you worry at the moment?"

Sleep — Sleep supports various functions of the human body such as repair and restorative processes [38] and is a crucial aspect for the well-being of an individual [39]. Prior research found that low levels of sleep can lead to lower self-esteem [40]. We hypothesized that "good" self-reported

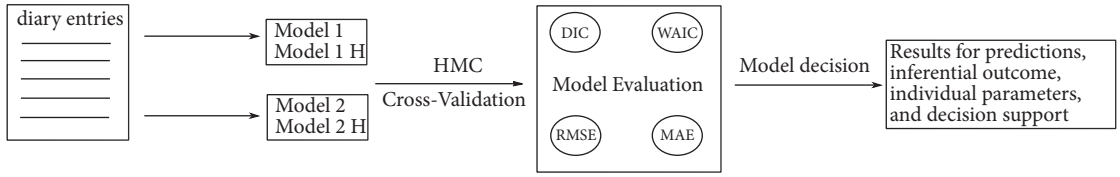


Figure 1: Graphic visualization of approach.

sleep levels can lead to higher levels of self-esteem. Sleep was assessed through the question "How well did you sleep last night?"

Enjoyed Activities — This concept relates to any action that has been executed by the participant that day. It describes to what degree the patient has relished a specific day by the performed activities. Since we assumed that joy—that in turn can trigger happiness—can potentially boost the self-esteem of individuals, we hypothesized that enjoyed activities are positively linked to self-esteem. The predictor enjoyed activities was assessed by the question "How much did you enjoy activities today?"

Social Contact — Social contact can provide important emotional support; and the lack thereof can be linked to depression [41]. We hypothesized a positive relationship between social contact and self-esteem. Social contact was assessed by asking the individuals "How much were you involved in social interaction today?"

2.2. Statistical Analysis

2.2.1. Approach. We applied two different models for predicting the self-esteem at time t based on the aforementioned predictors and their scores at time t, the ordered logit and stereotype logit model. Both approaches account for the ordinality in the measurements. Four models were eventually used because we modified each method by implementing patient-specific parameters in order to consider how they are individually affected by the psychological factors (Figure 1). We used Hamiltonian Monte Carlo techniques (HMC) for parameter estimation [42], applied cross-validation, and evaluated the models by comparing their outcomes based on various performance measures. We then utilized the model that performed best for illustrating the concrete predictions, the inferential outcomes (relationship between psychological factors and self-esteem), and the patient-individual parameters.

2.2.2. Ordinal Logistic Regression Model. One method that was utilized is the frequently used proportional odds or ordered logit model (OLM) that was initially proposed by McCullagh [25]. This model estimates the odds of observing a specific rank or less of self-esteem (score on the scale) for patient j at time step t for $rank = 1, \ldots, C$ where C is the number of ranks or the highest category on a scale (ten in our analysis since self-esteem is rated on a scale from one to ten) [43]:

$$\theta_{cjt} = \frac{P\left(rank \leq c \mid x_{jt}\right)}{P\left(rank > \mathrm{c} \mid x_{jt}\right)}. \tag{1}$$

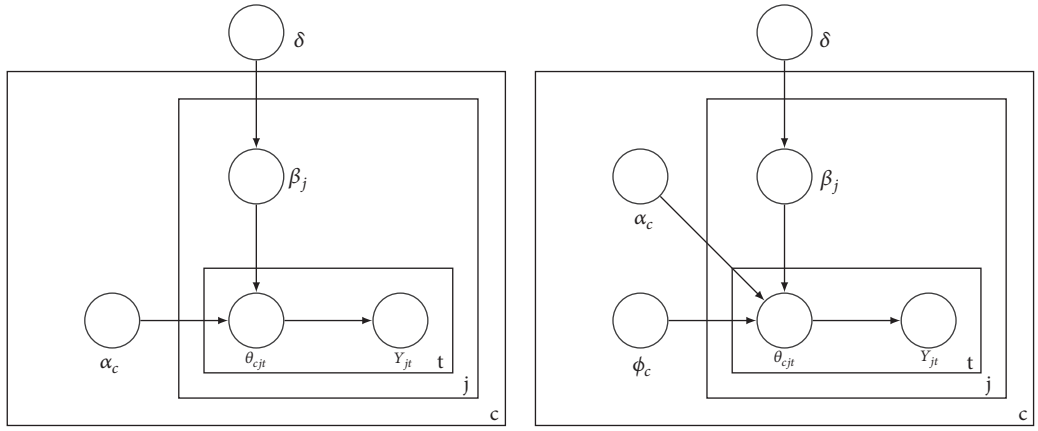
The estimation then follows (2). The parameters α_c are the boundaries of the categories or thresholds, also called cutpoints where $c = 1, \ldots, C-1$. This parameter has therefore nine distinct values. Furthermore, the cutpoints are following the constraint $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{C-1}$. x_{jt} is a vector of length five that represents the observations of the psychological factors for each patient j at each time step t. The β parameters are the weights to be estimated that reveal relationships between the factors and are utilized for the self-esteem prediction. This model is based on the proportional odds assumption. This means that the OLM assumes all β terms and their effects to be equal among all the levels of the dependent variable. As we can see, the β parameters do not vary among the ordinal levels or in any other fashion. Therefore, no individual effects are captured. The fixed β coefficient for all patients in the data leads to the unrealistic assumption that all individuals are similarly related to the psychological factors.

$$\ln\left(\theta_{cjt}\right) = \alpha_c - \left(x_{jt}\beta\right) \tag{2}$$

However, humans possess very unique and intricate qualities; each person has a different personality, opinion, thinking structure, and behavior; this can in turn lead to patient-individual effects from the predictors [44, 45]. We further assumed that including patient-individual parameters could lead to a greater prediction performance because more variance can potentially be explained. However, this process comes with a sacrifice of an increased model complexity. Nevertheless, we modified the model by introducing an additional index j into the β parameters which accounts for the varying effect a predictor can have on an individual. The OLM then yields the following form:

$$\ln\left(\theta_{cjt}\right) = \alpha_c - \left(x_{jt}\beta_j\right). \tag{3}$$

2.2.3. Stereotype Ordinal Logit Model. Another model that is less frequently used in research, presumably due to the rare existence of already implemented software packages [26, 46], is the stereotype ordinal logit model. This model was created by Anderson [27] in order to tackle the restrictive nature of the OLM due to its proportional odds assumption that is often violated in real datasets [47]. It can be seen as an extension of the multinomial logistic regression with the distinction that less parameters have to be estimated [46]. We additionally applied this model in order to compare the performance of both techniques and to demonstrate that heterogeneous parameters are not only beneficial when utilizing the OLM, but also in other statistical procedures. As in the OLM, θ_{cjt}



(a) Ordered logit model (b) Stereotype logit model

Figure 2: Graphic visualization for both models as plate notation.

is estimated; this is the odds of observing a specific rank of self-esteem in comparison to a baseline category (in our case the last category ten) for patient j at time t.

$$\theta_{cjt} = \frac{P\left(rank = c \mid x_{jt}\right)}{P\left(rank = \mathrm{C} \mid x_{jt}\right)} \tag{4}$$

The procedure of the stereotype logit model for the estimation is illustrated in (5) for $c = 1, \ldots, C$. As we can see by the index j, the β parameters already consider individual effects. The original model does not include this index. The $\alpha_c{}'s$ are the intercepts and the ϕ_c parameters are a score for the different levels of the outcome variable where $\alpha_1 = \phi_1 = 0$ [46]. Ordinality is only given as long as the constraint $0 = \phi_1 \leq \phi_2 \leq \cdots \leq \phi_C = 1$ is considered. Specifically, for a four-point scale, two $\phi's$ are to be estimated. For a ten-point scale, eight $\phi's$ are to be estimated.

$$P\left(rank = c \mid x_{jt}\right) = \frac{\exp\left(\alpha_c + \phi_c x_{jt}\beta_j\right)}{\sum_{c=1}^{C} \exp\left(\alpha_c + \phi_c x_{jt}\beta_j\right)} \tag{5}$$

2.2.4. Parameter Setting. Enabled by the Bayesian approach, we set different priors based on assumptions and already existing literature mentioned above. In this context, priors are beliefs in terms of probability distributions about the effects of the predictors that can be set before the actual data is considered. We set weak positive priors for the predictors mood, sleep, enjoyed activities, and social contact. For the variable worry, we set a weak negative prior. Implementing weak priors means sampling the corresponding parameter with high variance. Thereby, prior knowledge from related literature is taken into account while at the same time, the data strongly affects the analyses. Figure 2 illustrates the hierarchical structure of both models including heterogeneity parameters as a plate notation.

The parameters are distributed as shown in (6) where $\sigma^2 = 100$ (high variance). The expected value for the hyper-parameter δ is -1 or 1 depending on the definition as either a weak negative or positive prior. The parameters δ and $\alpha_{1\ldots C}$ are sampled from a normal distribution. The heterogeneous parameters for each patient, $\beta_{1\ldots J}$, are also sampled from a normal distribution; however, they are based on the vector δ. We decided to sample from a normal distribution because this allowed the parameters to evenly take on a positive or negative value. This means that we assumed that patients exist for whom a specific coefficient is positive whereas other patients are negatively affected. The results for δ indicate the effects each predictor has on the self-esteem on a population level. We utilized this parameter for prediction for the models that do not consider heterogeneity. The β parameters for each patient were used for the prediction of the individual models and illustration of the individual parameters.

$$\begin{aligned}
\delta &\sim \mathcal{N}\left(\mu \in \{-1, 1\}, \sigma^2\right) \\
\alpha_c &\sim \mathcal{N}\left(0, \sigma^2\right) \\
\gamma_c &\sim \mathrm{Dir}\,(\mathrm{A}) \\
\beta_j &\sim \mathcal{N}\left(\delta, \sigma^2\right) \\
Y_{jt} &\sim \mathrm{Cat}\left(\theta_{cjt}\right)
\end{aligned} \tag{6}$$

Solely in the stereotype model, as we can see in Figure 2, θ_{cjt} also depends on ϕ_c. This parameter is the cumulative sum of γ_c which follows a $Dirichlet(A_1, \ldots, A_C)$ distribution where $A_{1\ldots C} = 1$. Since the stereotype model requires ϕ_c to be steadily increasing, initialized with 0 and be limited to 1, sampling γ_c from a Dirichlet distribution is an appropriate procedure to meet this constraint [46]. As a final step, the actual predicted self-esteem level for each individual at each point in time (Y_{jt}) is sampled from a categorical distribution based on θ_{cjt}. For each model, we performed 60,000 iterations on four chains when running the Hamiltonian Monte Carlo algorithm and stored every twentieth draw from the last 30,000 iterations. We implemented the models in Python (https://www.python.org/) and utilized STAN [42] for Monte Carlo procedures.

Table 1: Results: performance for each model based on performance measures.

Model	RMSE	MAE	DIC	WAIC
No Heterogeneity Ordered logit	1.20	0.81	6204.88	6205.87
Heterogeneity Ordered logit	1.18	0.80	5914.91	6143.56
No Heterogeneity Stereo	1.21	0.82	6364.25	6369.72
Heterogeneity Stereo	1.08	0.73	5871.31	5772.14
Mean model	1.90	1.48	-	-
Mean individual model	1.38	0.98	-	-

2.2.5. Performance Measures. We implemented 10-fold stratified cross-validation in order to determine the model that achieves the best prediction performance. In 10-fold cross-validation, the dataset is divided into ten equally sized chunks (in our case each patient has observations in the training as well as the test dataset). Then, the models are trained on nine chunks and the tenth is predicted. This process is repeated ten times until every chunk is utilized as test data. 10-fold cross-validation is widely used and has also been shown to be suited for real-world datasets [48, 49].

We utilized the Deviance Information Criterion (DIC) [50] as indicator for measure of fit and model complexity [51]. The DIC is often used for model comparison and selection, especially in a Bayesian context [52]. The performance of a model is evaluated by the trade-off between how well the model fits the data and the complexity of the model. The model fit is expressed by the deviance (the lower the value, the better the fit), which is essentially the difference between a saturated model (a model that explains all variance in the responses) and the actual model. A penalty term is added to the model fit that is increasing with a rise in number of parameters [50]. Thus, models are preferred that have a smaller number of parameters. We chose the DIC as an indicator for model selection and comparison because it has been performing sufficiently regarding a variety of examples [51, 53].

According to Ando [54] and Richards and Richardson [55], however, the DIC can tend to prefer overfitted models and is only based on a point estimate [56, 57]. Thus, we also utilized the widely applicable or Watanabe-Akaike information criterion (WAIC) [58]. The WAIC is infrequently used in research and practice because of its additional computational effort [57]. According to Vehtari et al. [57], the WAIC represents an improvement of the DIC. Since the calculation for the number of parameters is based on each data point of the log likelihood, which is not the case for the DIC, the outcome is more stable and reliable. The WAIC (as well as the DIC) suggests a superior performance the smaller the value. For reasons of comparison and because of the mentioned issues regarding the DIC, we utilized both measures in our analyses. For readers interested in the exact derivations and steps regarding the calculation of the DIC and WAIC, we refer to the papers of Spiegelhalter et al. [50] and Vehtari et al. [57], respectively.

We further used the root-mean-square error (RMSE) and mean absolute error (MAE) as performance indicators. There is a debate about the selection of choosing either one of these measures. Willmott and Matsuura [59] and Willmott et al. [60], for example, criticized the usage of the RMSE and came to the conclusion that it is not a good indicator for the average model performance. They emphasized to only utilize the MAE since it is more natural compared to the RMSE. However, Chai and Draxler [61] showed that the RMSE can be a better indicator for model performance. Since there is no specific agreement in the literature as to which measure is more reliable, we decided to report both measures in our analysis.

Additionally, we defined a *mean model*. This model uses the arithmetic mean of the self-esteem value among the whole training set as prediction for each self-esteem value in the test data. Since we included heterogeneous parameters, we also used a *mean individual model* that utilizes the arithmetic mean of the training set on an individual patient level as predictions. We used these measures for comparison and as a baseline model; if we would not achieve a higher prediction performance than the *mean models*, it is questionable if the creation of such complex models is even worth the effort.

3. Results and Discussion

3.1. Principal Results. We can see that the *mean individual model* clearly performed better compared to the *mean model* (Table 1). It is also indicated that all created models performed better than the *mean models* regarding the RMSE and MAE (the other performance measures are not generatable for the *mean models*). Indicated by a Wilcoxon-Test, the errors differed significantly ($P < .05$). Therefore, creating such models is beneficial in regard to predictive performance in this context. The results further indicate that the implementation of patient-individual parameters was advantageous; both models performed better regarding each of the performance measures when accounting for individual effects even though the complexity of the models (number of parameters) increased (indicated by DIC as well as WAIC). This result highlights the importance of accounting for individual parameters. We can further see that the stereotype logit model benefit more from heterogeneity. Thus, we decided to utilize this model for further demonstration and analysis.

Figure 3 illustrates the predictive performance of this model in more detail. Specifically, it shows the observed values of self-esteem in the test data as a line and the predictions of the test data as crosses. The values are sorted in ascending order according to the observed values. Oftentimes, the predictions were the exact observed self-esteem value. Only

Table 2: Results: estimated model parameters including High Density Interval (significant parameters in bold).

Variables	Median	2.5% HDI	97.5% HDI
Mood	**16.82**	**14.25**	**19.55**
Worry	-1.05	−2.95	0.69
Sleep	1.50	-0.52	3.42
Enjoyed Activities	**4.26**	**2.37**	**6.24**
Social Contact	0.81	-1.34	2.82

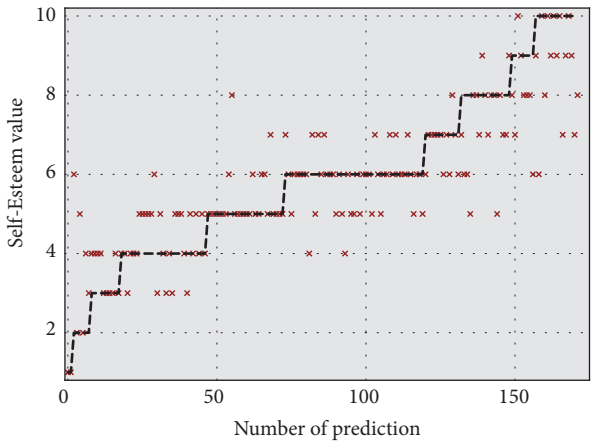


Figure 3: Graphic visualization of predicted and observed values.

once, the prediction was four categories off; however, it was frequently falsely predicted with a distance of two ranks. Since the predictions were close to the observed value most of the time, also indicated by the performance measures, we consider this a good result.

Table 2 demonstrates the effects of the psychological factors on the self-esteem. Here, the analysis was executed based on all data without withholding observations for evaluation of the models. The results indicate that the mood level of the patients is significantly related to the self-esteem.

Since recent literature found that low self-esteem is linked to depressive moods [62] and mood changes can modify self-concepts [63], this finding is plausible. As already indicated by Scheier et al. [64], who found that enjoyable leisure activities are related to factors for well-being, we show that enjoyed activities significantly increased the self-esteem. When individuals experience certain activities as fun and pleasure, they might be involved in actions that can boost their confidence, be of avail, and foster feelings of happiness that can in turn increase the sense of self-worth. Therefore, joy and doing well in a specific activity can potentially lead to feelings of reward and satisfaction and thus to an increased self-esteem.

The other predictors were not significant. However, for some of the patients, these predictors might be significantly related to the self-esteem. Figure 4 illustrates the distributions of the individual β parameters for each patient and each predictor. The values in this Figure cannot be read horizontally for each patient among the predictors; this means that the first patient for one predictor is not the same as the first patient for another predictor because the values are sorted in ascending order according to the individual mean value of the corresponding distribution. The horizontal line represents the zero value for the parameter and is an indicator for significance. The parameters varied tremendously, which again indicates the importance of considering heterogeneity. Even though the overall result for the variable worry, for instance, was insignificant, individuals exist for whom the outcome, the negative effect, was significantly true and vice versa. This finding occurs for every predictor except the mood level. Mood seemed to not be negatively related to self-esteem for any patient. Thus, the overall parameter for this predictor was highly significant. This individualized information can potentially help therapists to make refined and improved decisions on an individual level. Some patients were affected negatively by certain factors and some positively; with this procedure, it is possible to detect those specific patients. The gained information can lead to an increased understanding of patient-individual behavior and improved decision-making which can in turn result in personalized interventions and potentially better treatment outcomes.

3.2. Limitations. Besides the implications this study provides, we also depict some limitations and directions for further improvement and research opportunities. One limitation is the usage of diary data. Self-reported data is not inspected personally by a professional; even though this fact enables researchers to collect data in their natural environment, it lacks objectivity and can also lead to falsely reported data and social desirability bias [16, 65]. Furthermore, we measured self-esteem only based on one question. Even though this question is related to one item of the state self-esteem scale [29], it might not represent the whole complexity of self-esteem. We also obtained data for only 130 patients and 2326 observations. We believe that applying the modified models on other datasets in order to confirm the results can lead to an increased representativity. More data could improve the accuracy of gained information and especially enhance prediction performance. Therefore, more research in this context is necessary for a verification of the results.

Another aspect that can be viewed critically is the attempt of predicting a self-esteem value of a *new* patient that has not been seen before by the model. Unfortunately, even though we would have access to varying parameters for the individuals, we would not have any information on the *new* patient; therefore, we would predict the new patients' self-esteem based on the overall parameter δ. In fact, we would not perform less accurate compared to models that do not account for heterogeneous influences; however, we would also not benefit from the modified models. Nevertheless, after obtaining some information about the *new* patient and a recalculation of the models, we could obtain individual parameters for this patient. Thus, the utilization of the modified models is initially not beneficial for *new* patients, but after an initial data collection period, valuable results can be generated.

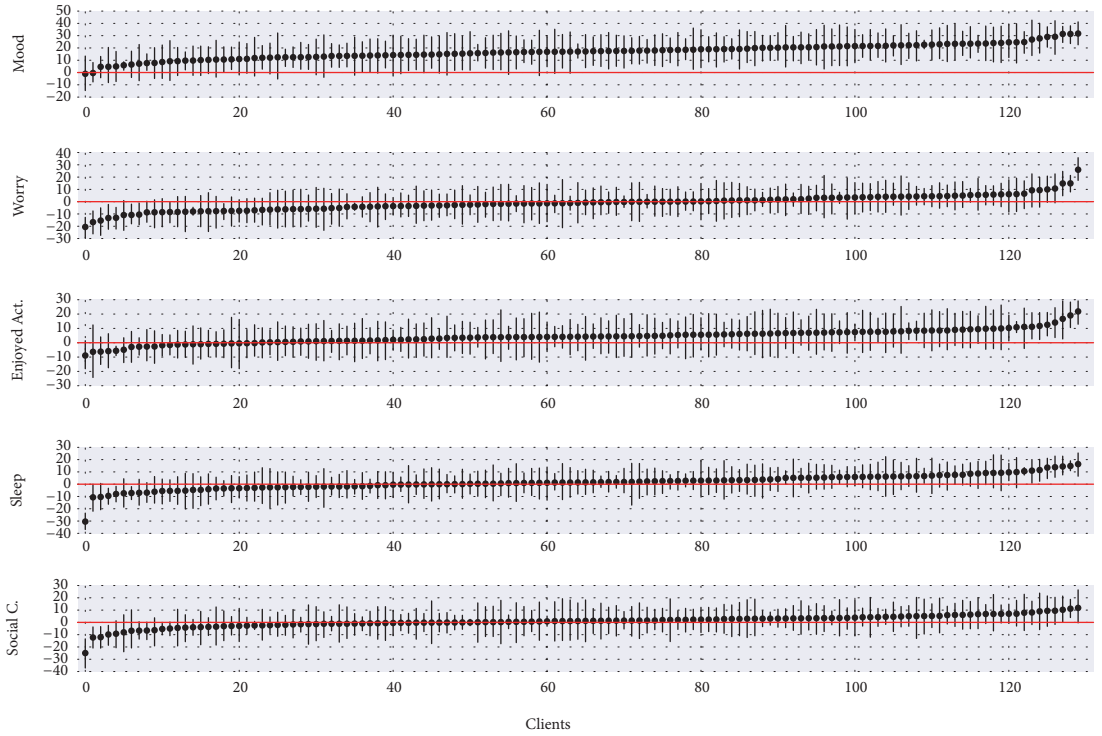


Figure 4: Graphic visualization of parameter distribution for each patient.

Another important aspect is the question of the exact impact of more accurate predictions. How can the illustrated improvement be translated into practical benefits? If a therapist is able to provide more refined recommendations, how are the individuals affected, how can this be converted into higher outcomes, and what role do costs play in this question? We seek to tackle challenges in this context in further research.

4. Conclusion

In this study, we predicted the self-esteem level of participants based on collected EMA data from a two-arm randomized control trial. We modified two statistical models by including heterogeneous slopes for each patient and employed Hamiltonian Monte Carlo techniques for parameter estimation. Therefore, one purpose of this study was to highlight the importance of individuality in such analyses. We illustrated a path of how individual parameters can be considered in an ordinal context and demonstrated how the prediction performance of different models is influenced by doing so. Individual parameters did not only increase the performance of these models but also allow practitioners to investigate differences among patients; possibly leading to knowledge gain and deeper insight about the patients. We further emphasized the importance of self-esteem in this context and investigated its relationships with other psychological factors. We found that the self-esteem level of patients was positively related to mood and when individuals experienced joyful activities. We further found that worries can be negatively linked to self-esteem whereas better sleep and social contact can be positively related to self-esteem. These latter results were not significant overall; however, we demonstrated that for some individuals these effects are significant. With our approach, we hope we can provide valuable information in the mental health sphere and support the decision-making process in personalized interventions.

Data Availability

The data used to support the findings of this study might be available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The current study has been conducted in the context of the EU FP7 Project E-COMPARED [Project no. 603098]. We therefore thank the EU for funding and the E-COMPARED consortium for the fantastic cooperation.

References

[1] S. Saddichha, M. Al-Desouki, A. Lamia, I. A. Linden, and M. Krausz, "Online interventions for depression and anxiety - a systematic review," *Health Psychology and Behavioral Medicine*, vol. 2, no. 1, pp. 841–881, 2014.

[2] I. Titzler, K. Saruhanjan, M. Berking, H. Riper, and D. D. Ebert, "Barriers and facilitators for the implementation of blended psychotherapy for depression: A qualitative pilot study of therapists' perspective," *Internet Interventions*, 2018.

[3] E. Karyotaki, H. Riper, J. Twisk et al., "Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms a meta-analysis of individual participant data," *JAMA Psychiatry*, vol. 74, no. 4, pp. 351–359, 2017.

[4] P. Carlbring, G. Andersson, P. Cuijpers, H. Riper, and E. Hedman-Lagerlöf, "Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis," *Cognitive Behaviour Therapy*, vol. 47, no. 1, pp. 1–18, 2018.

[5] A. W. Paradise and M. H. Kernis, "Self-esteem and psychological well-being: Implications of fragile self-esteem," *Journal of Social and Clinical Psychology*, vol. 21, no. 4, pp. 345–361, 2002.

[6] J. F. Sowislo and U. Orth, "Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies," *Psychological Bulletin*, vol. 139, no. 1, pp. 213–240, 2013.

[7] D. H. Silvera, T. D. Bergersen, L. Bjørgum, J. A. Perry, J. H. Rosenvinge, and A. Holte, "Analyzing the relation between self-esteem and eating disorders: differential effects of self-liking and self-competence," *Eating and Weight Disorders*, vol. 3, no. 2, pp. 95–99, 1998.

[8] K. H. Trzesniewski, M. B. Donnellan, T. E. Moffitt, R. W. Robins, R. Poulton, and A. Caspi, "Low self-esteem during adolescence predicts poor health, criminal behavior, and limited economic prospects during adulthood," *Developmental Psychology*, vol. 42, no. 2, pp. 381–390, 2006.

[9] G. Manna, G. Falgares, S. Ingoglia, M. R. Como, and S. D. Santis, "The relationship between self-esteem, depression and anxiety: Comparing vulnerability and scar model in the Italian context," *Mediterranean Journal of Clinical Psychology*, vol. 4, no. 3, pp. 1–17, 2016.

[10] A. E. Steiger, M. Allemand, R. W. Robins, and H. A. Fend, "Low and decreasing self-esteem during adolescence predict adult depression two decades later," *Journal of Personality and Social Psychology*, vol. 106, no. 2, pp. 325–338, 2014.

[11] A. E. Steiger, H. A. Fend, and M. Allemand, "Testing the vulnerability and scar models of self-esteem and depressive symptoms from adolescence to middle adulthood and across generations," *Developmental Psychology*, vol. 51, no. 2, pp. 236–247, 2015.

[12] A. Gustavsson, M. Svensson, F. Jacobi, C. Allgulander, J. Alonso, E. Beghi et al., "Cost of disorders of the brain in Europe 2010," *European Neuropsychopharmacology*, vol. 21, no. 10, pp. 718–779, 2011.

[13] D. Leger, "The cost of sleep-related accidents: A report for the National Commission on Sleep Disorders Research," *SLEEP*, vol. 17, no. 1, pp. 84–93, 1994.

[14] B. G. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson, "A systems approach to healthcare: Agent-based modeling, community mental health, and population well-being," *Artificial Intelligence in Medicine*, vol. 63, no. 2, pp. 61–71, 2015.

[15] M. Iida, P. E. Shrout, J-.P. Laurenceau, and N. Bolger, "Using diary methods in psychological research," in *APA Handb Res methods Psychol Vol 1 Found planning, Meas Psychom*, H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher, Eds., pp. 277–305, American Psychological Association, Washington, USA, 2012.

[16] D. S. Moskowitz and S. N. Young, "Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology," *Journal of Psychiatry & Neuroscience*, vol. 31, no. 1, pp. 13–20, 2006.

[17] W. Waegeman, B. De Baets, and L. Boullart, "ROC analysis in ordinal regression learning," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 1–9, 2008.

[18] J. B. Nezlek, "Using multilevel random coefficient modeling to analyze social interaction diary data," *Journal of Social and Personal Relationships*, vol. 20, no. 4, pp. 437–469, 2003.

[19] J. Long, "Regression models for nominal and ordinal outcomes," in *The SAGE Handbook of Regression Analysis and Causal Inference*, H. Best and C. Wolf, Eds., pp. 173–204, SAGE Publications, London, 2014.

[20] U. Jakobsson, "Statistical presentation and analysis of ordinal data in nursing research," *Scandinavian Journal of Caring Sciences*, vol. 18, no. 4, pp. 437–440, 2004.

[21] M. P. LaValley and D. T. Felson, "Statistical presentation and analysis of ordered categorical outcome data in rheumatology journals," *Arthritis & Rheumatism*, vol. 47, no. 3, pp. 255–259, 2002.

[22] M. Forrest and B. Andersen, "Ordinal scale and statistics in medical research," *British Medical Journal (Clinical Research ed.)*, vol. 292, no. 6519, pp. 537-538, 1986.

[23] D. Hedeker, "Methods for Multilevel Ordinal Data in Prevention Research," *Prevention Science*, vol. 16, no. 7, pp. 997–1006, 2015.

[24] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: capturing life as it is lived," *Annual Review of Psychology*, vol. 54, pp. 579–616, 2003.

[25] P. McCullagh, "Regression models for ordinal data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 42, no. 2, pp. 109–142, 1980.

[26] X. Liu, "Fitting stereotype logistic regression models for ordinal response variables in educational research (stata)," *Journal of Modern Applied Statistical Methods*, vol. 13, no. 2, pp. 528–543, 2014.

[27] J. A. Anderson, "Regression and ordered categorical variables," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 1, pp. 1–30, 1984.

[28] A. Kleiboer, J. Smit, J. Bosmans et al., "European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): Study protocol for a randomized controlled, non-inferiority trial in eight European countries," *Trials*, vol. 17, no. 1, p. 387, 2016.

[29] T. F. Heatherton and J. Polivy, "Development and Validation of a Scale for Measuring State Self-Esteem," *Journal of Personality and Social Psychology*, vol. 60, no. 6, pp. 895–910, 1991.

[30] Z. Graham, *Better Than a Stick in the Eye: A Method for Resolving Conflicts and Bringing about Changes in Marriages, Families, and the Workplace [Internet]*, AuthorHouse, 2009.

[31] P. C. Clasen, A. J. Fisher, and C. G. Beevers, "Mood-Reactive Self-Esteem and Depression Vulnerability: Person-Specific Symptom Dynamics via Smart Phone Assessment," *PLoS ONE*, vol. 10, no. 7, pp. 1–16, 2015.

[32] M. Rosenberg, *Society and the adolescent self-image*, Princeton University Press, Princeton, NJ, USA, 1965.

[33] R. W. Robins, H. M. Hendin, and K. H. Trzesniewski, "Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale," *Personality and Social Psychology Bulletin*, vol. 27, no. 2, pp. 151–161, 2001.

[34] S. Cohen and M. S. Rodriguez, "Pathways Linking Affective Disturbances and Physical Disorders," *Health Psychology*, vol. 14, no. 5, pp. 374–380, 1995.

[35] S. L. Minden, "Mood disorders in multiple sclerosis: diagnosis and treatment," *Journal of Neurovirology*, vol. 6, no. 2, pp. 160–167, 2000.

[36] J. Hoyer, E. S. Becker, and J. Margraf, "Generalized anxiety disorder and clinical worry episodes in young women," *Psychological Medicine*, vol. 32, no. 7, pp. 1227–1237, 2002.

[37] G. J. Diefenbach, M. E. McCarthy-Larzelere, D. A. Williamson, A. Mathews, G. M. Manguno-Mire, and B. G. Bentz, "Anxiety, depression, and the content of worries," *Depression and Anxiety*, vol. 14, no. 4, pp. 247–250, 2001.

[38] G. Curcio, M. Ferrara, and L. De Gennaro, "Sleep loss, learning capacity and academic performance," *Sleep Medicine Reviews*, vol. 10, no. 5, pp. 323–337, 2006.

[39] E. K. Gray and D. Watson, "General and specific traits of personality and their relation to sleep and academic performance," *Journal of Personality*, vol. 70, no. 2, pp. 177–206, 2002.

[40] S. Lemola, K. Räikkönen, V. Gomez, and M. Allemand, "Optimism and self-esteem are related to sleep. Results from a large community-based sample," *International Journal of Behavioral Medicine*, vol. 20, no. 4, pp. 567–571, 2013.

[41] N. Frasure-Smith, F. Lespérance, G. Gravel et al., "Social support, depression, and mortality during the first year after myocardial infarction," *Circulation*, vol. 101, no. 16, pp. 1919–1924, 2000.

[42] B. Carpenter, A. Gelman, M. D. Hoffman et al., "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, vol. 76, no. 1, pp. 1–32, 2017.

[43] M. J. Norusis, "Ordinal Regression," in *PASW Stat 180 Adv Stat Proced Companion*, pp. 69–89, Prentice Hall, 2010.

[44] S. L. Gable, H. T. Reis, and A. J. Elliot, "Behavioral activation and inhibition in everyday life," *Journal of Personality and Social Psychology*, vol. 78, no. 6, pp. 1135–1149, 2000.

[45] S. M. Weinstein and R. Mermelstein, "Relations between daily activities and adolescent mood: The role of autonomy," *Journal of Clinical Child & Adolescent Psychology*, vol. 36, no. 2, pp. 182–194, 2007.

[46] J. Ahn, B. Mukherjee, M. Banerjee, and K. A. Cooney, "Bayesian inference for the stereotype regression model: application to a case-control study of prostate cancer," *Statistics in Medicine*, vol. 29, no. 25, pp. 997–1003, 2012.

[47] X. Liu and H. Koirala, "Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data," *Journal of Modern Applied Statistical Methods*, vol. 11, no. 1, pp. 242–254, 2012.

[48] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, vol. 14 of *IJCAI'95*, pp. 1137–1143, 1995.

[49] G. McLachlan, K. A. Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data [Internet]*, Wiley, 2005.

[50] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–616, 2002.

[51] A. Berg, R. Meyer, and J. Yu, "Deviance information criterion for comparing stochastic volatility models," *Journal of Business & Economic Statistics*, vol. 22, no. 1, pp. 107–120, 2004.

[52] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for Bayesian models," *Statistics and Computing*, vol. 24, no. 6, pp. 997–1016, 2014.

[53] D. Spiegelhalter, N. G. Best, and B. P. Carlin, *Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models*, 1998.

[54] T. Ando, "BAYesian predictive information criterion for the evaluation of hierarchical BAYesian and empirical BAYes models," *Biometrika*, vol. 94, no. 2, pp. 443–458, 2007.

[55] D. Richards and T. Richardson, "Computer-based psychological treatments for depression: A systematic review and meta-analysis," *Clinical Psychology Review*, vol. 32, no. 4, pp. 329–342, 2012.

[56] M. Plummer, "Penalized loss functions for Bayesian model comparison," *Biostatistics*, vol. 9, no. 3, pp. 523–539, 2008.

[57] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, pp. 1–20, 2016.

[58] S. Watanabe, "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory," *Journal of Machine Learning Research (JMLR)*, vol. 11, pp. 3571–3594, 2010.

[59] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.

[60] C. J. Willmott, K. Matsuura, and S. M. Robeson, "Ambiguities inherent in sums-of-squares-based error statistics," *Atmospheric Environment*, vol. 43, no. 3, pp. 749–752, 2009.

[61] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[62] P. Martyn-Nemeth, S. Penckofer, M. Gulanick, B. Velsor-Friedrich, and F. B. Bryant, "The relationships among self-esteem, stress, coping, eating behavior, and depressive mood in adolescents," *Research in Nursing & Health*, vol. 32, no. 1, pp. 96–109, 2009.

[63] D. A. DeSteno and P. Salovey, "The Effects of Mood on the Structure of the Self-concept," *Cognition & Emotion*, vol. 11, no. 4, pp. 351–372, 1997.

[64] S. D. Pressman, K. A. Matthews, S. Cohen et al., "Association of enjoyable leisure activities with psychological and physical well-being," *Psychosomatic Medicine*, vol. 71, no. 7, pp. 725–732, 2009.

[65] D. E. Logan, R. L. Claar, and L. Scharff, "Social desirability response bias and self-report of psychological distress in pediatric chronic pain patients," *PAIN*, vol. 136, no. 3, pp. 366–372, 2008.