Baryshnikov, Ilya; Aledavood, Talayeh; Rosenström, Tom; Heikkilä, Roope; Darst, Richard; Riihimäki, Kirsi; Saleva, Outi; Ekelund, Jesper; Isometsä, Erkki

Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: A prospective ecological momentary assessment study on 80 psychiatric outpatients

Published in:
Journal of Affective Disorders

DOI:

Published: 01/03/2023

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

Published under the following license:
CC BY

Please cite the original version:
Short communication

Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: A prospective ecological momentary assessment study on 80 psychiatric outpatients

Ilya Baryshnikov, Talayeh Aledavood, Tom Rosenström, Roope Heikkilä, Richard Darst, Kirsi Riihimäki, Outi Saleva, Jesper Ekelund, Erkki Isometsä

* Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
* Department of Computer Science, Aalto University, Espoo, Finland
* Division of Mental Health and Substance Abuse Services, Department of Health and Social Services, Helsinki, Finland
* Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

ARTICLE INFO

Keywords:
Ecological momentary assessment
Patient Health Questionnaire (PHQ-9)
Depressive symptoms

ABSTRACT

Background: Depression-related negative bias in emotional processing and memory may bias accuracy of recall of temporally distal symptoms. We tested the hypothesis that when responding to the Patient Health Questionnaire (PHQ-9) the responses reflect more accurately temporally proximal than distal mood states.

Methods: Currently, depressed psychiatric outpatients (N = 80) with depression confirmed in semi-structured interviews had the Aware application installed on their smartphones for ecological momentary assessment (EMA). The severity of “low mood”, “hopelessness”, “low energy”, “anhedonia”, and “wish to die” was assessed on a Likert scale five times daily during a 12-day period, and thereafter, the PHQ-9 questionnaire was completed. We used auto- and cross-correlation analyses and linear mixed-effects multilevel models (LMM) to investigate the effect of time lag on the association between EMA of depression symptoms and the PHQ-9.

Results: Autocorrelations of the EMA of depressive symptom severity at two subsequent days were strong (r varying from 0.7 to 0.9; p < 0.001). “Low mood” was the least and “wish to die” the most temporally stable symptom. The correlations between EMA of depressive symptoms and total scores of the PHQ-9 were temporally stable (r from 0.3 to 0.6; p < 0.001). No effect of assessment time on the association between EMA data and the PHQ-9 emerged in the LMM.

Limitations: Altogether 11.5% of observations were missing.

Conclusions: Despite fluctuations in severity of some of the depressive symptoms, patients with depression accurately recollect their most dominant symptoms, without a significant recall bias favouring the most recent days, when responding to the PHQ-9.

1. Introduction

Measurement-based care of depression (Trivedi et al., 2006), currently a standard in clinical practice, includes a systematic assessment of the severity of depressive symptoms and guides treatment-related decision-making. The Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001) is a widely used validated self-report depression scale that scores each of the nine Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) (American Psychiatric Association, 2013) criteria over the preceding two weeks. However, current mood may, by resulting in autobiographic memory bias (Hitchcock et al., 2019; Quinlivan et al., 2017), distort recall of distal emotional experiences in patients with mood disorders. Therefore, responses to the PHQ-9 might reflect more patients’ temporally proximal than distal depressive symptoms within the two-week period.

Ecological momentary assessment (EMA), a method involving a frequent repeated assessment of patients’ experiences, symptoms, and behaviour in real time and in the context of their daily life, may overcome the recall bias related to use of depression scales. Studies examining how daily EMA of depression symptoms corresponds to the total

Received 29 November 2021; Received in revised form 21 November 2022; Accepted 23 December 2022
Available online 28 December 2022
0165-0327/© 2022 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/).
score of PHQ-9 are, however, rare and include small (Cao et al., 2020; Moukaddam et al., 2019; Torous et al., 2015) or non-clinical samples with self-reported depressive symptoms (Burchert et al., 2021; McIntyre et al., 2021). These studies have demonstrated moderate correlations between PHQ-9 scores and daily assessments of depression symptoms and a within-patient fluctuation of some of the depressive symptoms. For example, mood symptoms (Bowen et al., 2017) and suicidal ideation (Kleiman et al., 2017) seem to oscillate over time in depressed patients. This temporal fluctuation of symptom severity of depression, in addition to recall bias, may also affect the validity of the PHQ-9, supposed to reflect the most frequent and dominant depression symptoms over a two-week period, as required in the DSM-5 for major depressive disorder (MDD).

In this prospective study using EMA in psychiatric patients with major depressive episodes, we aimed to examine the temporal patterns of core depression symptoms ("low mood", "anhedonia", "low energy", "hopelessness", and "wish to die") severity assessed daily over the course of two weeks. We also investigated the effect of assessment timing on the association between the EMA of the daily depression symptom severity and patients' retrospective self-report of their depression symptoms over the two weeks, assessed by PHQ-9. We hypothesized that a) depressive symptom severity fluctuates over time, when assessed daily; b) the PHQ-9 scores reflect more the severity of depressive symptoms during temporally proximal days than distal days.

2. Methods

2.1. MoMo-Mood project

This study is a part of the Mobile Monitoring of Mood (MoMo-Mood) study—a collaborative project between the Department of Psychiatry, University of Helsinki, the Department of Psychiatry, Helsinki University Hospital, the Department of Psychiatry, University of Turku, the Department of Psychiatry, Turku University Hospital, the City of Espoo Mental Health Services, and the Department of Computer Science at Aalto University. The Helsinki and Uusimaa Hospital District's Ethics Committee approved the research protocol on 29 August 2018.

2.2. Participants

Overall, 80 outpatients with DSM-5 (American Psychiatric Association, 2013) diagnoses of MDD (n = 46); currently depressed patients with bipolar disorder (BD) (n = 16), or borderline personality disorder (BPD) (n = 18) were recruited from outpatient clinics for mood disorders. Inclusion criteria were a) MDD, BD type I or II, or BPD with an ongoing major depressive episode; b) PHQ-9 score ≥ 10; and c) possession of a smartphone with an Android or iOS operating system. MDD and BD were ascertained with the Mini International Neuropsychiatric Interview (Sheehan et al., 1998) and BPD with the Structured Clinical Interview-II (First et al., 1997). Exclusion criteria were a) psychotic features; b) concurrent substance use disorder; and c) imminent risk of suicide.

2.3. Data collection

We used Aalto University's Niima Data Collection Platform (Aalda-wood et al., 2017). The Niima is primarily developed to run studies with various modalities of behavioural and physiological data, enabling various types of data from study participants to be collected and assessed. It allows different levels of access to data (e.g. access only to certain sources of data) for different researchers, depending on their role in the study. The platform is privacy-first; and does not use other cloud services as data intermediaries. Niima is an open-source project that also allows use of other open-source applications as necessary. In MoMo-Mood, we used AWARE application as part of Niima for collecting the EMA data.

2.4. Procedures

All participants after giving informed consent had the AWARE application (Ferreira et al., 2015) installed on their smartphones. During the two-week period patients received notifications on their phones 5 times during the day (once in the morning, one in the evening, and three times in between). The patients were asked to assess their mood, ability to feel pleasure, hopefulness, energy level, and willingness to live using Likert scale items (1–7). The mood item of the Likert scale varied from "I am very happy" – 1 to "I am very depressed" – 7; the energy level item from "I am totally exhausted" – 1 to "I am very energetic" – 7; the pleasure item "I am enthusiastic and feel pleasure" (1 – "not at all", 7 – "very much"); the hopefulness item from "I am totally hopeless" – 1 to "I am very hopeful" – 7; and the willingness to live item from "I want to die" – 1 to "I want to live" – 7. The patients fulfilled the PHQ-9 both before and the day after the EMA. Due to technical reasons the data on the first two days of the EMA was abnormal in some patients. Therefore, we marked the day of the PHQ-9 after the EMA as the day 0 and tracked the EMA responses back to the day-12.

2.5. Statistical analysis

In the analyses, we calculated means of the daily EMA item scores using all available data and investigated product-moment auto- and cross-correlations. No imputation was used. We estimated linear multilevel models (Bates et al., 2015; Gelman and Hill, 2007), regressing the experience sampling variables on the fixed-effects of lag-1 experience (autocorrelation), post-active PHQ-9 report, time lag to last sampling, and interaction of PHQ-9 and the lag to last sampling. The interaction represents the hypothesis that PHQ-9 is more correlated with the temporally proximal than with the distal experiences. As a sensitivity analysis, we compared a few recent days with the entire 12 days using a dummy indicator of time lag from 0 to 2 days from the last sampling (and PHQ-9 sampling).

3. Results

3.1. Sample characteristics

The sample comprised 61 women and 19 men (mean age 35 years; s. d. 12 years). The mean PHQ-9 scores were 16.5 (SD 5.2) before and 16.1 (SD 5.6) after the EMA (NS). Overall, 11.5 % of all EMA but no PHQ-9 data were missing (all available data were used, and no patient was fully excluded from any analysis). The daily EMA means are presented in the Table 1.

3.2. Correlation analyses

The experience sampling variables of each day were correlated with the variable itself at the last assessment date, although slightly more so during the immediately preceding days (Fig. 1a). The mood variable appeared less stable over time than death thoughts (Fig. 1b). The correlation of the experience sampling variables (and their daily total scores) with post-active PHQ-9 reports remained much more stable over the 12-day follow-up than their autocorrelations (Fig. 1c vs. a). Confidence intervals of successive days clearly overlapped when examining the correlation of core symptoms with post-active PHQ-9 (Fig. 1d).

3.3. Linear multilevel regression analysis

In the linear multilevel regression analysis, experience samples were associated with their past values and with PHQ-9, but not with time of experience assessment relative to PHQ-9 (see Supplementary Table 1). In addition, as a sensitivity analysis, we estimated the same models without the lag-1 variable for autocorrelations, but the conclusion of the lack of assessment time-to-PHQ interactions on experience samples
remained (see Supplementary Table 2). The conclusion also withstood a test of non-linear association of the few closest days to PHQ-9 assessment with experiences (see Supplementary Table 3).

4. Discussion

In this prospective ecological momentary assessment (EMA) study on 80 psychiatric outpatients with depression, we tested the hypothesis that PHQ-9 reflects more accurately temporally proximal rather than distal depressive symptoms due to previously postulated memory- and mood-related negative bias. Overall, the PHQ-9 does not seem to favour the severity of depressive symptoms during the days that are more proximal to the responding, partly owing to relative temporal stability of the majority of depressive symptoms over the 12-day period, except for

<table>
<thead>
<tr>
<th>EMA items</th>
<th>Time from the last report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood</td>
<td>Day-11 Day-10 Day-9 Day-8 Day-7 Day-6 Day-5 Day-4 Day-3 Day-2 Day-1 Day 0</td>
</tr>
<tr>
<td></td>
<td>(1.18) (1.3) (1.1) (1.2) (1.3) (1.2) (1.2) (1.1) (1.3) (1.2) (1.0) (1.2)</td>
</tr>
<tr>
<td>Hopelessness</td>
<td>(1.12) (1.21) (1.25) (1.21) (1.44) (1.24) (1.27) (1.29) (1.28) (1.21)</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>(1.27) (1.28) (1.29) (1.34) (1.29) (1.42) (1.31) (1.38) (1.31) (1.22)</td>
</tr>
<tr>
<td>Death</td>
<td>(1.53) (1.71) (1.74) (1.78) (1.62) (1.58) (1.76) (1.75) (1.74) (1.78)</td>
</tr>
<tr>
<td>Energy</td>
<td>(1.08) (1.19) (1.17) (1.02) (1.26) (1.17) (1.21) (1.19) (1.02)</td>
</tr>
</tbody>
</table>

Explanations: The “mood” item on the 1–7 Likert scale varied from “I am very happy” – 1 to “I am very depressed” – 7; the “hopelessness” item from “I am totally hopeless” – 1 to “I am very hopeful” – 7; the “anhedonia” item was “I am enthusiastic and feel pleasure” (1 – “not at all”, 7 – “very much”); “death” item from “I want to die” – 1 to “I want to live” – 7; and the “energy” item from “I am totally exhausted” – 1 to “I am very energetic” – 7.
moderate temporal stability of the majority of depressive symptoms over a two-week period when answering the PHQ-9.

In conclusion, despite variability in the severity of some depressive symptoms and previously postulated negative memory bias, patients with depression are able to recollect their most dominant mood state over a 12-day period.

CRediT authorship contribution statement

Ilya Baryshnikov, MD, PhD, has recruited patients and control subjects, participated in designing this study, conducted literature searches and drafted the manuscript as first author.

Talayeh Aledavood, PhD, is co-PI of the MoMo-Mood project, and has participated in designing this study and been responsible for smartphone applications and data management.

Tom Rosenstrom, PhD, has participated in designing this study and conducted statistical analyses.

Roopo Heikkinen, MSc, has been responsible for clinical data and smartphone collection.

Richard Darst, PhD, has been responsible for smartphone applications, database and data security in the MoMo-Mood project.

Kiri Rihinäki, MD, PhD, has recruited patients and control subjects.

Ottis Saleva, MD, has recruited patients and control subjects.

Jesper Ekelund, MD, PhD, has participated in designing the study.

Erkki Isometsä, MD, PhD, is the PI of the MoMo-Mood project. He supervised study design, analyses and drafting of the manuscript.

All authors have critically read and provided expert comments on the manuscript plus approved the final version.

Role of the funding sources

This study is part of the Mobile Monitoring of Mood (MoMo-Mood) research project, which has been funded by a grant (TYH2021312) from the Helsinki and Uusimaa Hospital District (HUS) with EJ as the PI. TA acknowledges the support by James S. McDonnell Foundation through a postdoctoral fellowship in “Understanding Dynamic & Multi-Scale Systems”.

Nobody besides the researchers had a role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2022.12.127.

References


