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

Baryshnikov, Ilya

2023-03-01

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 [a], Talayeh Aledavood [b], Tom Rosenström [a,d], Roope Heikkilä [a], Richard Darst [b], Kirsi Riihimäki [c], Outi Saleva [a], Jesper Ekelund [a], Erkki Isometsä [a,*]

[a] Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[b] Department of Computer Science, Aalto University, Espoo, Finland
[c] Division of Mental Health and Substance Abuse Services, Department of Health and Social Services, Helsinki, Finland
[d] Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

A B S T R A C T

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

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 (Aledavood 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 accessed. 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

Table 1

The daily means and standard deviations (SD) of the ecological momentary assessment (EMA) questions during a 12-day follow-up of patients with depression (*n* = 80).

EMA items[a]	Time from the last report											
	Day-11	Day-10	Day-9	Day-8	Day-7	Day-6	Day-5	Day-4	Day-3	Day-2	Day-1	Day 0
Mood	3.78 (1.18)	3.8 (1.3)	3.84 (1.1)	4.05 (1.31)	4.06 (1.28)	4.02 (1.06)	3.93 (1.3)	3.8 (1.2)	3.93 (1.36)	4.02 (1.3)	3.97 (1.29)	3.91 (1.25)
Hopelessness	4.15 (1.12)	3.99 (1.21)	3.92 (1.18)	3.75 (1.25)	3.72 (1.21)	3.83 (1.1)	3.94 (1.44)	3.96 (1.24)	3.85 (1.27)	3.77 (1.29)	3.79 (1.28)	3.85 (1.21)
Anhedonia	4.08 (1.27)	4.12 (1.28)	4.19 (1.18)	4.28 (1.29)	4.33 (1.34)	4.27 (1.29)	4.27 (1.42)	4.16 (1.31)	4.19 (1.4)	4.25 (1.38)	4.23 (1.31)	4.21 (1.22)
Death	5.47 (1.53)	5.4 (1.63)	5.35 (1.71)	5.27 (1.74)	5.23 (1.78)	5.38 (1.62)	5.34 (1.8)	5.51 (1.58)	5.35 (1.76)	5.29 (1.75)	5.23 (1.74)	5.22 (1.78)
Energy	3.57 (1.08)	3.6 (1.18)	3.48 (1.1)	3.37 (1.19)	3.39 (1.17)	3.35 (1.02)	3.39 (1.26)	3.57 (1.17)	3.44 (1.21)	3.36 (1.19)	3.37 (1.02)	3.4 (1.15)

[a] 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.
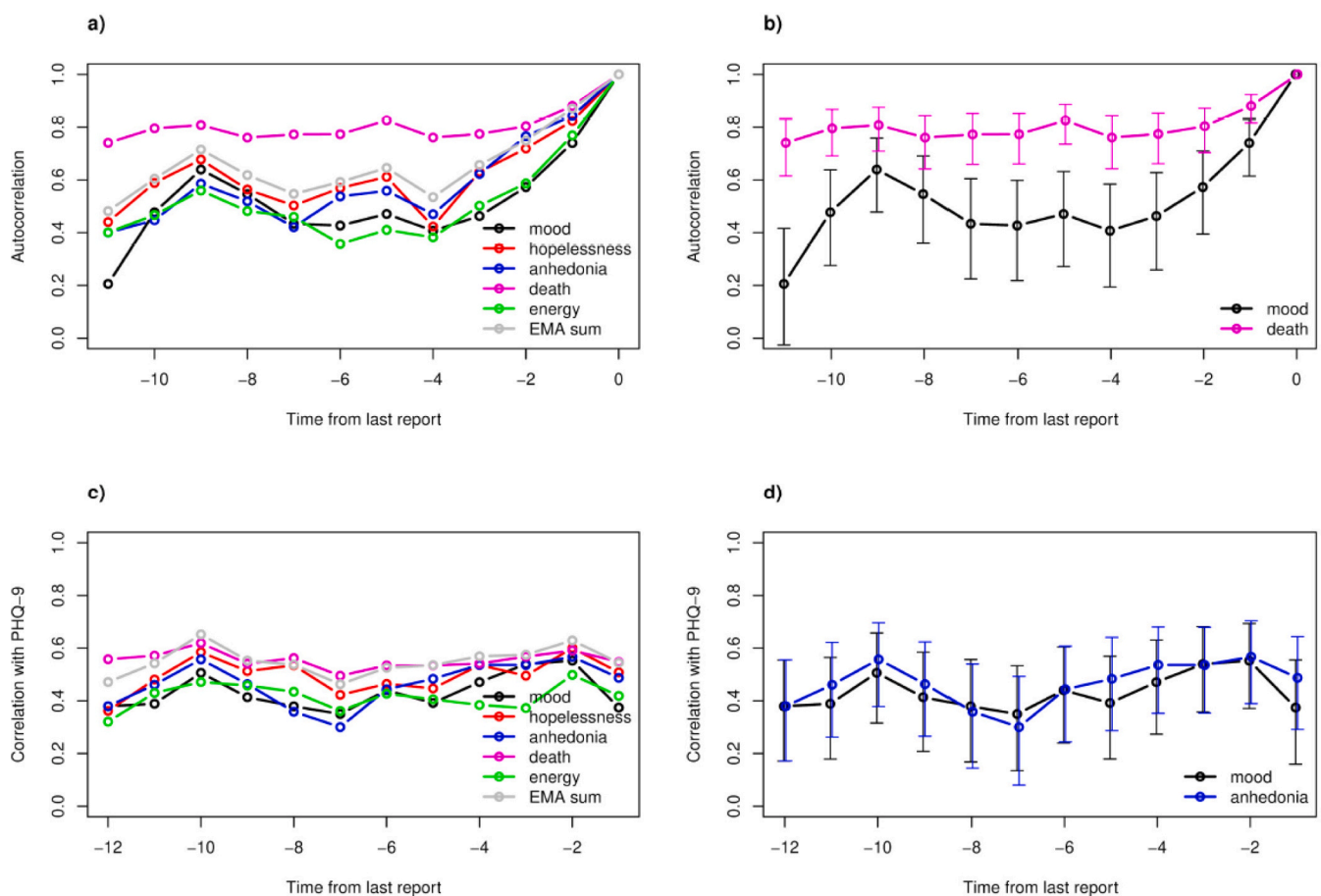

Fig. 1. a) Autocorrelation of daily averages of experience sampling variables with the last day's value. b) Mood and death thoughts variables from panel (a) with whiskers for 95 % confidence intervals. c) The (cross-)correlation of experience sampling variables with the post-active PHQ-9 sum score. d) The values of panel (c) core symptoms are shown with 95 % confidence intervals. In panels (a) and (b), day–11 is the first day with EMA data and day 0 is the last day with EMA data. In panels (c) and (d) day 0 is the day when the PHQ-9 was answered and days-1, -2,…, -12 are the 12 days prior to the PHQ-9 where the EMA was completed. Abbreviations: EMA – ecological momentary assessment; PHQ-9 – Patient Health Questionnaire.

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

depressive mood, which markedly fluctuated in some patients with depression.

This study has several strengths. First, to our knowledge, this is the largest EMA study examining the association between daily EMA of depression symptoms and PHQ-9 in psychiatric outpatients with depression ($N = 80$). Second, our study included also currently depressed patients with BD and BPD, allowing broader generalizability of the findings. Third, mood disorders were diagnosed with semi-structured clinical interviews by experienced psychiatrists.

Our study also has limitations. First, not all depressive symptoms were included in our EMA items. Second, 11.5 % of the EMA data were missing and no imputation method was used. Third, we were unable to gather EMA data on depressive symptoms during the full two-week period that the PHQ-9 requires (fell two days short). Fourth, we deliberately used mean values of five daily assessments in the analyses, which limits our ability to unravel within-day symptom severity fluctuations. Fifth, the EMA items were not validated against the healthy controls. However, correlations between the EMA-responses and PHQ-9 scores support convergent validity of the EMA items. Sixth, effects of treatment were not investigated. Finally, all patients were diagnosed with an ongoing major depressive episode, limiting the range of their PHQ-9 scores and mood fluctuations.

The EMA of daily rated depressive symptoms allows examination of their temporal patterns. In our study, although varying over a 12-day period, autocorrelations of such important depressive symptoms as "wish to die", "hopelessness", "low energy", and "anhedonia" remained high or moderate. This is in contrast to depressive mood, which showed notable temporal variability in some patients, with the autocorrelations weakening with increasing time lag between assessments. This finding corresponds to the previous EMA studies in patients with MDD (Thompson et al., 2012; Torous et al., 2015), revealing instability of mood in some depressed patients. Fluctuating mood is probably even more distinguished in depressed patients with comorbid BDP (Trull et al., 2008), patients also included in our sample. In contrast, the EMA symptom "wish to die" appeared the most temporally stable, differing from a report describing rapid temporal fluctuation of suicidal ideation in depressed patients (Kleiman et al., 2017). Repeated answers of the patients with absent "wish to die" (low base rate) as well as conceptual differences between "wish to die" and suicidal thoughts may underlie this discrepancy. Overall, although varying, severity of the majority of core depressive symptoms appears moderately stable, apart from a mood symptom showing the most prominent fluctuation in some depressed patients over a 12-day period.

The correlations between the EMA of depressive symptoms (and their daily total scores) and the total scores of the PHQ-9 were stable and moderate throughout the 12-day period, agreeing with previous research (Burchert et al., 2021; McIntyre et al., 2021; Moukaddam et al., 2019). Moreover, we detected no associations between timing of assessment of depressive symptom severity and PHQ-9 scores. The relative temporal stability of correlations between the EMA of depressive symptoms and the PHQ-9 scores concur with our observation of moderate temporal stability of the majority of depressive symptoms over the 12 days. The variability in severity of some of the depressive symptoms, detected in the EMA, does not appear sufficiently marked to bias patients' ability to recall their most frequent and dominant depressive symptoms over the two weeks when answering the PHQ-9, which inquired about the frequency of depressive symptoms over this period.

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 two-week period when answering the PHQ-9.

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

Roope Heikkilä, 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.

Kirsi Riihimäki, MD, PhD, has recruited patients and control subjects.

Outi 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 off 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.

Conflict of interest

The authors declare no conflict of interest.

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

Aledavood, T., Triana Hoyos, A.M., Alakörkkö, T., Kaski, K., Saramäki, J., Isometsä, E., Darst, R.K., 2017. Data collection for mental health studies through digital platforms: requirements and Design of a Prototype. JMIR Res. Protoc. 6, e110.

American Psychiatric Association, 2013. The diagnostic and statistical manual of mental disorders, Fifth edition. Arlington, VA, American Psychiatric Association.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bowen, R., Peters, E., Marwaha, S., Baetz, M., Balbuena, L., 2017. Moods in clinical depression are more unstable than severe normal sadness. Front. Psychiatry 8, 56.

Burchert, S., Kerber, A., Zimmermann, J., Knaevelsrud, C., 2021. Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: comparison with the PHQ-9 depression screening. PLoS One 16, e0244955.

Cao, J., Truong, A.L., Banu, S., Shah, A.A., Sabharwal, A., Moukaddam, N., 2020. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. JMIR Ment. Health. 7, e14045.

Ferreira, D., Kostakos, V., Dey, A.K., 2015. AWARE: mobile context instrumentation framework. Front. ICT 2, 6. https://doi.org/10.3389/fict.2015.00.

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., Benjamin, L.S., 1997. Structured clinical interview for DSM-IV axis II personality disorders, (SCID-II). American Psychiatric Association, Washington, DC.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Hitchcock, C., Rodrigues, E., Rees, C., Gormley, S., Dritschel, B., Dalgleish, T., 2019. Misremembrance of things past: depression is associated with difficulties in the recollection of both specific and categoric autobiographical memories. Clin. Psychol. Sci. 7, 693–700.

Kleiman, E.M., Turner, B.J., Fedor, S., Beale, E.E., Huffman, J.C., Nock, M.K., 2017. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J. Abnorm. Psychol. 126, 726–738.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.

McIntyre, R.S., Lee, Y., Rong, C., Rosenblat, J.D., Brietzke, E., Pan, Z., Park, C., Subramaniapillai, M., Ragguett, R., Mansur, R.B., Lui, L.M.W., Nasri, F., Gill, H., Berriah, S., 2021. Ecological momentary assessment of depressive symptoms using the mind.me application: convergence with the patient health questionnaire-9 (PHQ-9). J. Psychiatr. Res. 135, 311–317.

Moukaddam, N., Truong, A., Cao, J., Shah, A., Sabharwal, A., 2019a. Findings from a trial of the smartphone and OnLine usage-based eValuation for depression (SOLVD) application: what do apps really tell us about patients with Depression? Concordance between app-generated data and standard psychiatric questionnaires for depression and anxiety. J. Psychiatr. Pract. 25, 365–373.

Quinlivan, E., Dallacker, M., Renneberg, B., Strasser, E., Fiebig, J., Stamm, T.J., 2017. Autobiographical memory in bipolar disorder and its link to neuropsychological functioning. Psychopathology 50, 246–254.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 22–33 quiz 34–57.

Thompson, R.J., Mata, J., Jaeggi, S.M., Buschkuehl, 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.

Torous, J., Staples, P., Shanahan, M., Lin, C., Peck, P., Keshavan, M., Onnela, J.P., 2015. Utilizing a personal smartphone custom app to assess the patient health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment. Health. 2, e8.

Trivedi, M.H., Rush, A.J., Wisniewski, S.R., Nierenberg, A.A., Warden, D., Ritz, L., Norquist, G., Howland, R.H., Lebowitz, B., McGrath, P.J., Shores-Wilson, K., Biggs, M.M., Balasubramani, G.K., Fava, M., STAR*D Study Team, 2006. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am. J. Psychiatry 163, 28–40.

Trull, T.J., Solhan, M.B., Tragesser, S.L., Jahng, S., Wood, P.K., Piasecki, T.M., Watson, D., 2008. Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. J. Abnorm. Psychol. 117, 647–661.