THE ACCEPTABILITY OF REAL-TIME HEALTH MONITORING AMONG COMMUNITY PARTICIPANTS WITH DEPRESSION: A SYSTEMATIC REVIEW AND META-ANALYSIS OF THE LITERATURE

Giovanni de Girolamo[1], Chiara Barattieri di San Pietro[2], Viola Bulgari[3], Jessica Dagani[1], Clarissa Ferrari[4], Matthew Hotopf[5], Giuseppe Iannone[1], Ambra Macis[4], Faith Matcham[5], Inez Myin-Germeys[6], Aki Rintala[6], Sara Simblett[5], Til Wykes[7], Cristina Zarbo[1] and on behalf of the RADAR-CNS Consortium[5]

[1] Psychiatric Epidemiology and Evaluation Unit, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[2] Department of Psychology, University of Milano-Bicocca, Milan, Italy

[3] IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[4] Statistical Unit, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[5] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[6] Department for Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium

[7] Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

RUNNING TITLE: ESM/EMA acceptability for patients with MDD

Corresponding Author:

Giovanni de Girolamo, M.D.

IRCCS Istituto Centro San Giovanni di Dio, Fatebenefratelli

Via Pilastroni, 4 – 25125 Brescia (Italy)

Email: gdegirolamo@fatebenefratelli.eu

ABSTRACT

Background: The application of Experience Sampling Method/Ecological Momentary Assessment (ESM/EMA) methods to individuals with Major Depressive Disorder (MDD) seems promising, but evidence about their acceptability is still unclear. The aim of this systematic review and meta-analysis (registration number CRD42017060438) was to investigate the acceptability of ESM/EMA techniques for health monitoring in patients with MDD, by examining the dropout rate and related-reasons, and to explore the effects of individual, methodological, and technical features on dropping out.

Method: According to PRISMA guidelines, after leading a systematic search on major electronic databases, a structured process for selecting and collecting data was followed.

Results: A total of 19 studies were included in the analyses. From results, it emerged a dropout rate of 3.6%. Our findings show that the use of paper and pencil tools in combination with electronic devices, the time-based sampling method and not providing monetary incentives significantly increase the dropout rate of patients with MDD during ESM/EMA assessment. Age, gender, depression severity, duration of monitoring, number of assessments each day and number of questions did not affect dropout rate.

Conclusions: Results of this systematic review may assist clinicians and researchers in planning, implementing or evaluating the use of ESM/EMA to assess the health status of community-based individuals with MDD.

KEYWORDS: dropout, community, Ecological Momentary Assessment, Experience Sampling Method, depression, acceptability

INTRODUCTION

In the last decades, advances in technology have seen the application of electronic devices to healthcare settings, enhancing the quality of the assessment, monitoring and treatment of a range of different populations in ecological settings (Mohr, Zhang, & Schueller, 2017). In psychiatry, the Experience Sampling Method Method (ESM; Csikszentmihalyi & Larson, 2014; Myin-Germeys et al., 2018) and the Ecological Momentary Assessment (EMA; Stone & Shiffman, 1994) have made possible to repeatedly monitor and assess psychological, psychophysiological, biological, and behavioral data embedded in the context of daily life (Trull & Ebner-Priemer, 2013), by means of a range of different tools (e.g. electronic or paper-pencil diaries, sensors, smartphone applications, software, actigraphs, mobile phones, etc.). These methods are promising for their application with individuals with mental disorders, as they have the advantage of focusing on real-life conditions and provide a better picture of patients' experience in their natural environments (i.e. assessing and monitoring moods, thoughts, symptoms, and behaviors and their fluctuation over time, as well as characteristics of the environment, like location, time of the day, etc.), thus minimizing recall biases and providing ecologically valid data (Myin-Germeys, Klippel, Steinhart, & Reininghaus, 2016; Shiffman, Stone, & Hufford, 2008; Trull & Ebner-Priemer, 2009).

ESM/EMA techniques have been widely used with patients with a diagnosis of Major Depressive Disorder (MDD), with the aim to better understand diurnal mood variations, reactivity to stress, sleep, Positive Affect (PA) and Negative Affect (NA) during the daily life (Armey, Schatten, Haradhvala, & Miller, 2015; Ebner-Priemer & Trull, 2009; Myin-Germeys et al., 2009; Telford, McCarthy-Jones, Corcoran, & Rowse, 2012). Indeed, due to MDD clinical features (e.g. intense daily emotional fluctuation, high emotional reactivity to events, recall biases, etc.), the use of ESM/EMA in a real-time environment may provide many data useful for both researchers and clinicians, even in terms of prediction of long-term depression severity (Peeters, Berkhof, Rottenberg, & Nicolson, 2010). The adoption of EMA by patients with depression has increased our knowledge of clinical characteristics of depression (mainly in terms of emotion reactivity, cortisol patterns, or daily rumination; Colombo et al., 2019), of the impact of treatment on daily life, of residual symptoms in remitted patients, and of the clinical features of depression in pediatric populations (aan het Rot, Hogenelst, & Schoevers, 2012).

Nevertheless, acceptability of ESM/EMA in patients with MDD requires a specific attention as these techniques may induce high reactivity (like increased psychological burden, awareness of mental state, or fatigue) and may be perceived as too intrusive or overwhelming (particularly due to typical characteristics of MDD as described by the DSM-5, like chronic loss of interest or pleasure, high emotional reactivity, fatigue, and drop of some cognitive abilities; American Psychiatric Association, 2013). To this regards, in a qualitative study on community samples across three European Countries (i.e., Italy, Spain, and the UK), we found that the engagement of individuals with MDD with remote measurement was commonly affected by motivation, including perceived burden and inconvenience related to the ESM/EMA measurements (Simblett et al., 2019).

Two recent systematic reviews on the acceptability of ESM/EMA (in terms of adherence to daily prompts) for individuals with different mental disorders (Vachon, Viechtbauer, Rintala, & Myin-Germeys, 2019) and individuals with MDD (Colombo et al., 2019) found high levels of adherence (i.e. higher than 70%). However, the systematic review of Torous et al. (2019)on adults with depressive symptoms, found a dropout rate of 26.2%. Acceptability of ESM/EMA seems to be related to individuals' sociodemographic and clinical characteristics (i.e. gender, age, substance use, psychiatric disorder) as well as to methodological features (i.e. assessment time within the day, incentives, sampling scheme, number of evaluations per day, human feedbacks, in-app mood monitoring) and study characteristics (i.e. sample size) (Messiah, Grondin, & Encrenaz, 2011; Rintala, Wampers, Myin-Germeys, & Viechtbauer, 2019; Torous et al., 2019; Vachon et al., 2019).

However, despite knowledge on the acceptability of ESM/EMA is increasing, evidence on patients with MDD is still unclear and requires a specific attention. Therefore, the main aim of this systematic review and meta-analysis was to systematically explore the acceptability of ESM/EMA techniques of outpatients with MDD living in the community. In particular, we aimed to:

1. examine the dropout rate and related-reasons with ESM/EMA techniques used to assess the health status of individuals with MDD;

2. examine whether the acceptability of ESM/EMA differs in relation to participants' individual and clinical characteristics as well as to the characteristics of the techniques employed.

METHODS

Protocol and registration

In accordance with the Preferred Reporting Items for Systematic review and Meta-analysis (PRISMA) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009), our systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 04 April 2017 and was last updated on 12th March 2019 (registration number CRD42017060438).

Eligibility criteria

Based on data available in published papers, we defined the following eligibility criteria: (i) population: adult individuals (age ranging from 18 to 65 years old) with a current diagnosis of MDD made with a structured diagnostic interview (e.g. SCID, MINI) by a specialized mental health professional, and living in the community; (ii) intervention: ESM/EMA techniques used to assess and monitor the subject's health status over time; (iii) comparators: no restrictions on comparison groups; (iv) study design: original articles, no restrictions on design; (v) clear indication of the number of individuals with MDD who dropped out during ESM/EMA monitoring. No date limits were imposed. We excluded: (i) studies including individuals with remitted depression, dysthymia, with mixed mental disorders, without reporting separate data for individuals with MDD, or with severe mental disorders in comorbidity (i.e. substance abuse disorders, psychotic disorders, and bipolar disorders); (ii) studies which used screening tools (like PH9, BDI, CES-D) to establish a diagnosis of MDD; (iii) case reports, dissertation, protocols, reviews, case-series studies, unpublished studies and studies in languages other than English; (iv) studies using ESM/EMA for other interventions than monitoring (e.g., Cognitive-Behavioural Therapy) or for passive monitoring only (e.g., actigraphy).

Information sources and search strategy

All published peer-reviewed articles were retrieved through a systematic literature search on Web of Science, PubMed and PsycINFO databases from inception to 12th March 2019. The final search strategy was*: ("active monitoring" OR "experience sampling method" OR "active remote monitoring" OR smartphone OR app OR "ecological momentary assessment" OR "daily diar*" OR "electronic diar*"OR "computer-assisted diar*"OR "ambulatory assessment*" OR "electronic momentary assessment" OR "hand-held computer" OR "structured diar*" OR "mobile") AND (depress* OR "mood disorder*" OR "depress* symptom*")*. Furthermore, the reference lists of relevant articles and reviews were hand-searched to locate additional studies not identified by electronic searches.

Study selection

To ensure consistency across reviewers, a calibration exercise was accomplished prior to the formal screening. Five reviewers (GI, CB, VB, JD, CZ) independently screened the article titles and abstracts against the inclusion and exclusion criteria. Then, four authors (GI, CZ, CB and VB) performed independent systematic reviews and data extraction from all full-text articles and three reviewers (GI, JD, CZ) extracted again data from all the full-text articles for cross-checking. Any disagreements in the data extraction process were negotiated among all reviewers. Where two or more studies reported overlapping samples, priority was given to the study with the largest sample size.

Data Collection Process

Data from the included studies were extracted independently by five researchers (GI, CB, VB, CZ and JD) according to a data extraction form developed by two authors, which worked independently, and stored in a customized online structured spreadsheet. When available, data were extracted for the following items: (1) study characteristics: authorship, year, country of recruitment, study design, study primary aim; (2) sample characteristics: number of participants with MDD who entered the study, age mean and standard deviation of individuals with MDD, gender of individuals with MDD, depression severity; (3) ESM/EMA characteristics: duration of ESM/EMA monitoring (in days), sampling method (i.e., time based, event based, or both), type of device used (i.e., both electronic and paper and pencil; electronic

only; paper and pencil only), number of daily sampling (e.g. alerts, calls); number of items each day (i.e. questions to which the individual is required to answer each day); receiving monetary incentive upon completion of the assessment (categorized as yes or no); (4) outcome measures: number of individuals with MDD who dropped the monitoring with ESM/EMA and related percentage in relation to the number of participants who started the ESM/EMA monitoring; (5) reasons for dropping: qualitative motivations for dropping the ESM/EMA monitoring.

Outcome Measures

The acceptability of ESM/EMA techniques used to assess the health status of people with MDD in daily life settings was the primary outcome of our review. In line with a recent theoretical model (Sekhon, Cartwright, & Francis, 2017), we defined acceptability in terms of ratio between the number of subjects who dropped the study and the total number of study participants at baseline. Whenever the original study comprised both an observational phase (usually, pre-treatment) and a treatment phase using ESM/EMA, we considered only those participants who dropped out during the observational phase (i.e. during ESM/EMA monitoring), so as to exclude any confounding factors due to the effect of treatment.

In line with the International Conference on Harmonization (ICH) E9 guideline, "dropout" was intended as "*a subject in a clinical trial who for any reason fails to continue in the trial until the last visit required of him/her by the study protocol*" (ICH Expert Working Group, 1998, p. 33). Specifically, for ESM/EMA studies, we considered a participant as dropping out when, according to the study authors, he/she was excluded from the analysis as he/she failed to complete the ESM/EMA assessments. To this regards, it should be noted that in different studies researchers have employed various criteria to exclude participants from the analysis because of dropout, and there is not an official 'quantitative' definition of dropout: some have used a minimum of 30% of filled-out reports as the threshold for being left in the sample and be considered completers, whereas others have used a 50 or even an 80% inclusion rate.

For secondary outcomes, we examined whether the provision of monetary incentive, the ESM/EMA characteristics (i.e., type of device, sampling method, duration of monitoring, number of daily sampling, and number of daily items), as well as participants characteristics

(i.e., age, gender, level of depression severity at the beginning of the study), might influence the acceptability of ESM/EMA techniques among community participants with MDD.

Methodological Quality Assessment and Risk of Bias

Methodological quality assessment of the included studies was done independently by four researchers (GI, CZ, CB and VB) using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (National Heart Lung and Blood Institute, 2018), which allows to establish the presence (i.e., "yes") or absence (i.e., "no") of certain factors as possible biases in a study. According to original procedures, the overall methodological quality was judged as "good", "fair", or "poor", where "good" indicates the least risk of bias, "fair" suggests that there are some bias not sufficient to invalidate results, and "poor" denotes a significant risk of bias.

Statistical Analysis

A descriptive analysis of the characteristics of studies, subjects and ESM/EMA techniques was performed, emphasizing similarities and differences between studies. A meta-analysis was performed through a random-effects model to consider the variability (heterogeneity) of the studies included in the meta-analysis (Viechtbauer, 2005). A pooled prevalence estimate for dropout rate of subjects with MDD was calculated via an ad-hoc methodological option to properly consider the zero frequencies/proportions (Viechtbauer, 2010) by adding an arbitrarily small positive quantity (equal to 0.00025 that resulted in a good trade-off between reliability and feasibility) to the zero frequencies, in line with a common methodological procedure and without losses in generalization (Efthimiou, 2018).

Mixed-effects meta-regression models were carried out to measure the relationship between the dropout rate of MDD participants and selected continuous and discrete variables used as moderators. In details, a subgroup analysis was performed for categorical moderators and standard meta-regression models were performed for the continuous moderators. All data were analyzed using the "metaphor" package (Viechtbauer, 2010) of R software (R Core Team, 2015 http://www.R-project.org/).

RESULTS

Study selection

Starting from 9,800 records identified through the database (N=7,951) and other sources (N=1,849), the systematic search resulted in 5,697 records after removing duplicates (Figure 1). During the title and abstract screening process, we excluded 5,098 articles for the following reasons: they did not include depression (e.g., irrelevant studies), were written in a language other than English, had no full-text available or were papers from the same study (e.g., overlapping samples, multiple publications). When titles and abstracts appeared to meet the inclusion criteria or when there was any uncertainty, the articles were kept in for the full-text screening process. Overall, 599 full-text articles were screened for eligibility: 580 of them were excluded mainly because they included individuals with no established current diagnosis of MDD (or no reporting separate data for MDD group, or inpatients with MDD, or patients with MDD in comorbidity with other severe mental disorders). Nineteen studies were included in the systematic review. An assessment of inter-rater reliability was made by calculating agreement on 10% of the screened titles and abstract articles (Cohen's k: 0.92) and on 10% of the examined full texts (Cohen's k: 0.87).

Figure 1 here

Studies Characteristics

Table 1 reports the main characteristics of studies included in the meta-analysis. Data on age (mean and SD), gender, depression severity of individuals with MDD have been reported differently in each study. Indeed, in some studies the data referred to the number of participants with MDD who started the EMA/ESM assessment – including patients who dropped the study-, while in other studies they referred to those who concluded the entire study. Due to the main aim of this systematic review, we included data only of those studies that referred to patients with MDD which started the EMA/ESM assessment and monitoring.

Table 1 here

Sample Demographics

Studies varied in their sample sizes (from a low of 4 up to 76 participants with MDD, with a mean number of participants in each study of 34) and included a total of 639 participants with MDD, most commonly adult females. Gender, mean age and standard deviation of

participants with MDD who started the ESM/EMA monitoring were reported only in 8 studies. Four studies did not reported data about the mean age and standard deviation of the participants, while 7 studies reported the data referring to the number of participants who completed the entire study. Mean age ranged from 26 to 48 years old.

Clinical Characteristics of MDD Participants

All participants lived in the community and had a diagnosis of MDD according to specified diagnostic criteria established by validated diagnostic instruments (e.g. SCID, MINI). The level of depression severity was extracted based on the mean reported in the article. Three studies did not report data about the patients' level of depression, while 9 reported data about the number of participants who completed the entire study and were then not included in the final analyses. Depression severity, reported in 7 studies, ranged from mild to severe (3 studies included subjects with severe depression).

Characteristics of ESM/EMA methods

ESM/EMA monitoring lasted generally 7 days, although there was a marked variability in assessment duration, ranging from 1 day up to 150 days. The frequency of daily sampling generally was 8 times per day (range 1-31 per day). Each day, each individual answered to in mean 90 questions/items (range 9-330 questions per day).

Overall, most studies followed a time-based sampling method (N=15, 78.9%), 4 studies (21.0%) used both event-based and time-sampling methods. Electronic devices (either alone or in combination with paper and pencil instruments) were used in 18 out of 19 studies; since only one study used only paper and pencil, that category was not used for the analyses. Finally, about half studies (N=10; 52.6%) provided monetary incentives for participants who completed the study.

Measures of Acceptability

Participants' dropout was observed in 11 studies (61.1%), while in the remaining studies participants completed the ESM/EMA monitoring as foreseen and there were no dropouts

(Table 1). Dropout ranged from 0% to 16.7%, with a pooled dropout rate of 3.6% [95% CI: 1.4; 5.9]: this was significantly different from zero (p=0.002) (Figure 2).

Figure 2 here

Reasons for dropping out were reported in 6 studies (31.6%). The main reasons for dropping out were related to compliance issues, technical and individual difficulties related to ESM/EMA use, voluntary dropout, adverse personal event during monitoring, and researcher time constraints. Only one study gave information about the personal experience of participants using ESM/EMA techniques (Barge-Schaapveld, Nicolson, van der Hoop, & DeVries, 1995), and reported that dropout participants judged the study "annoying".

Associations with Dropout Rate

We included depression severity, sampling method, type of devices, and monetary incentive as possible moderators in the subgroup analysis (Table 2). Dropout prevalence in participants assessed with time-based sampling method was 5.2% (p=0.001), whereas dropout prevalence in participants assessed with event-based or event- and time-based sampling methods was not significantly different from 0 (p=0.917).

Dropout rate among participants using only an electronic device was equal to 0.04% (p=0.673), whereas dropout among participants using both electronic and paper and pencil devices was 9.7% (p=0.048).

Among participants who did not receive any monetary incentive prevalence dropout was 4.5% (p=0.028), as opposed to a dropout rate among those who received monetary incentive equal to 2.2% (p=0.049).

Table 2 here

Results of the meta-regression analysis (univariate)

Age, sex, duration of the assessment, number of daily samplings and number of daily items were included in different meta-regression models (Table 3). None of these variables was significantly associated with dropout rate (all p's>0.05).

Table 3 here

Methodological Quality and Risk of Bias within studies

The methodological quality of the 19 studies was generally of medium-high standard, being rated as good or fair in 18. A detailed description of quality assessment of each study is shown in Table 4.

Table 4 here

DISCUSSION

In this meta-analysis, we examined the acceptability of ESM/EMA assessment and monitoring of the health status of individuals with major depressive disorder living in the community. In line with the definition of acceptability proposed by Sekhon et al. (2017), we were interested in the number of participants who dropped out from the study, and we also looked at possible correlates of dropping out.

The overall dropout rate of individuals with a diagnosis of MDD during ESM/EMA monitoring was 3.6%. This data is markedly lower of the dropout rate found by a recent systematic review by Torous et al. (2019) on the acceptability of ESM/EMA for adults with depressive symptoms: they found an overall dropout rate of 26.2%. This difference might be due to the different studies included in the two systematic reviews. Indeed, the systematic review of Torous et al. (2019) included RCT studies which used smartphone apps to provide a psychological intervention for depressive symptoms, and the sample included both clinical and non-clinical populations. On the contrary, our systematic review included studies which used ESM/EMA only to assess and monitor the overall health status in clinical populations with MDD. We may hypothesize that patients with MDD (recruited mainly in clinical settings) who used ESM/EMA only for monitoring (vs treatment) are more likely to be compliant with the use of ESM/EMA. Indeed, recruitment of participants in clinical settings and their engagement in the study may be supported by treating clinicians. Moreover, using ESM/EMA to monitor rather than to treat may be less demanding in terms of time, personal resources and individual defenses.

Sociodemographic and clinical factors

Our review found no dropout differences based on demographic and clinical characteristics of the sample (e.g., age, gender, depressive symptomatology severity). Interestingly, contrary to our preliminary hypothesis, depression severity was not related to the likelihood to drop studies with ESM/EMA monitoring. We speculated that other personal factors, besides depression severity, may help shed some light on this result. For instance, the rate of relapses is an important variable worth investigating: it may be that, regardless of depressive symptoms' severity, individuals who experience frequent but shorter relapses may be more likely to dropout compared to individuals who experience longer, but less frequent relapses. Cognitive (e.g. difficulty concentrating, distractibility, forgetfulness, reduced reaction time, reduced brain processing speed, etc.), but also psychological (e.g. apathy, abulia, etc.) and physical (e.g. lack of energy, disturbed sleep, fatigue, etc.) symptoms of depression may all be related to dropout rates, especially if participants are required to answer questions which require cognitive and psychological efforts. In addition, other individual factors, such as motivation, feelings of self-worth, and suicide risk might be specifically associated with dropout rate. In addition, no effect of age (means, SD) was found for dropout rate. This finding may be related to the fact that the participants were in mean all in middle-adult age (30th-40th years old) and confident with new technologies (used in 95% of studies included).

Factors related to ESM/EMA techniques and the study

Factors related to the duration of the monitoring, the number of daily samplings and the number of daily items were not related to drop-out rate during ESM/EMA monitoring, disconfirming the assumption that a greater burden in daily life assessments might represent a possible hurdle to the engagement of patients with MDD. On the contrary, the modality of the sampling based on time (vs both based on event and time), the combination of both paper/pencil and electronic instruments (vs the use of only electronic device) and the absence of monetary incentive for completing the study, are related to an increased likelihood of dropout in the ESM/EMA monitoring.

In particular, our findings suggest that a higher dropout rate was associated with time-based sampling method during daily life (i.e. the individual has to answer to some questions at prefixed times during the day), rather than both time and event-based sampling method.

This result suggests that educating the participant to collect autonomously data in specific moments of the day (e.g., after wakening) is more effective to improve compliance than at only fixed times. Indeed, prefixed times are not personalized on the specific needs and habits of each individual, and sometimes alarms/prompts may be overlooked or perceived as annoying. To this regard, a recent study (van Genugten et al., 2020) explored the reasons for missing ESM/EMA assessments and found that most responders declared to 'being busy with an activity' and 'being asleep' (respectively 57.2% and 21.2%) while received the prompt for the daily evaluation. Furthermore, giving autonomy to participants may allow them to acquire much more self-efficacy, engagement and motivation in relation to the study and the ESM/EMA assessment.

In addition, our results show that the risk of dropping out was almost equal to 0 in participants using electronic devices only, but when participants used a combination of electronic devices and paper and pencil dropout rate increased to 9.7%. It might be plausible to assume that using only electronic tools, which may be easier to manage, faster and smarter than paper and pencil, is going to be preferable and increases the likelihood of participation. The use of electronic devices takes advantages for both researchers and participants as allows to collect different data in only one device, providing also information storage, feedbacks and secure data transfer.

Finally, we found that receiving no monetary incentive was associated with a dropout rate of 4.5% (vs 2.2% for studies which provided this incentive). Basing on the principles of operant conditioning and behavioral economics (Skinner, 1965), the monetary incentive may be classified as a positive reinforcement (i.e. a stimulus presented in response to a desired behavior which increases the likelihood of that behavior). In this case, monetary incentives provided at the end of the study acted as positive reinforcement of compliance with ESM/EMA monitoring. Monetary incentives may play a key role particularly for in individuals with MDD as, due to the pathology, their work life may be compromised (in terms of reduced performance, number of absences at work, and sometime also dismissal or resignation) and monetary difficulties may arise, resulting in higher motivation in being compliant with the study to gain the incentive.

Clinical implications of the findings

The results of this systematic review and meta-analysis have a range of clinical implications for both clinicians and researchers. Our findings show that compliance of patients with MDD for ESM/EMA assessment is not influenced by the severity of the depressive symptomatology or by the number of "stressors" provided each day. Therefore, also patients with severe symptomatology may be effectively engaged in ESM/EMA monitoring, over and above the amount of assessments.

On the contrary, clinicians and researchers who want to use ESM/EMA with this population should focus on the importance of a careful training for participants before starting the assessment, in order to increase their engagement, motivation, self-efficacy, as well as reduce technical and individual difficulties with electronic devices. Furthermore, the acceptability of ESM/EMA assessment may benefit from periodic feedbacks and support for possible emerging difficulties (technical, cognitive or affective), and from the use of personalized assessment times in combination with prefixed times. Being regularly in contact with participants and giving more feedback during the study might facilitate participants' level of engagement and reduce dropout.

Clinicians and researchers should also prefer the use of electronic device instead of paper and pencil instruments, as they allow to collect greater amount of data in a simpler and faster manner with advantages for both professionals and participants. Furthermore, electronic reminders, notifications and prompts may help to overcome dropout risk due to cognitive deficits common of MDD.

Finally, professionals should provide incentive/positive reinforcements to patients in order to strengthen their engagement with the daily monitoring. In clinical setting, this may be applied providing periodic positive verbal feedbacks to participants about the efforts they are doing to be compliant with the procedure and positive changes.

CONCLUSION

The results of this systematic review and meta-analysis suggest that the dropout rate for ESM/EMA studies in community participants with clinical depression is 3.6%. The use of

paper and pencil in combination with electronic devices, only time-based sampling method and not providing monetary incentives can significantly increase patients' dropout rate.

When interpreting the results, one should consider the limitation of this review. Studies that directly compared acceptability between individuals with and without depression were very limited in number, thus hindering the possibility to draw any strong conclusions on possible between-group differences. Likewise, almost all studies did not report any clinical and/or individual characteristics of the participants who dropped out, making it impossible to compare their characteristics with those of subjects who did not dropout. Another limitation is the lack of information on completeness of data collected during ESM/EMA. Furthermore, due to a limited number of studies reporting adherence rates (i.e. the number of answered prompts; 6/19) and their heterogeneity, it was not possible to investigate how this factor may be associated with demographic, clinical or ESM/EMA method variables.

Given the clinical relevance of data collected using ESM/EMA data among community samples for healthcare professionals, it may help if future research will focus also on the role played by comorbidities in affecting their acceptance, and start analyzing people's subjective experiences, taking these findings in due consideration

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

CONFLICT OF INTERESTS STATEMENT

The authors do not have any conflicts of interest to declare.

CONTRIBUTIONS

GDG, GI, VB, CZ and CB wrote the manuscript. GI, VB, JD, CB, CZ contributed to the development of the selection criteria, assessment strategy and data extraction. VB, CB, AR, GI, SS, FM developed the search strategy. CF and AM conducted the statistical analysis. FM/SS provided linguistic editing. All authors read, provided feedback and approved the final manuscript.

FUNDING

The RADAR-CNS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA (www.imi.europa.eu). This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

ACKNOWLEDGEMENTS

We thank our colleagues from the RADAR-CNS consortium and the involved institutions for their contribution to the development of this protocol.

DISCLAIMER

This communication reflects the views of the RADAR-CNS consortium and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

REFERENCES

aan het Rot, M., Hogenelst, K., & Schoevers, R. A. (2012). Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. Clinical psychology review, 32(6), 510-523.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. Washington DC: APA.

Armey, M. F., Schatten, H. T., Haradhvala, N., & Miller, I. W. (2015). Ecological momentary assessment (EMA) of depression-related phenomena. Current opinion in psychology, 4, 21-25.

Barge-Schaapveld, D. Q., Nicolson, N. A., van der Hoop, R. G., & DeVries, M. W. (1995). Changes in daily life experience associated with clinical improvement in depression. Journal of Affective Disorders, 34(2), 139-154.

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. Cognition and Emotion, 23(5), 1021-1040.

Cohen, L. H., Gunthert, K. C., Butler, A. C., Parrish, B. P., Wenze, S. J., & Beck, J. S. (2008). Negative affective spillover from daily events predicts early response to cognitive therapy for depression. Journal of Consulting and Clinical Psychology, 76(6), 955.

Colombo, D., Fernández-Álvarez, J., Patané, A., Semonella, M., Kwiatkowska, M., García-Palacios, A., . . . Botella, C. (2019). Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. Journal of clinical medicine, 8(4), 465.

Conrad, A., Wilhelm, F. H., Roth, W. T., Spiegel, D., & Taylor, C. B. (2008). Circadian affective, cardiopulmonary, and cortisol variability in depressed and nondepressed individuals at risk for cardiovascular disease. Journal of psychiatric research, 42(9), 769-777.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method Flow and the foundations of positive psychology (pp. 35-54): Springer.

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. Psychological assessment, 21(4), 463.

Efthimiou, O. (2018). Practical guide to the meta-analysis of rare events. Evidence-based mental health, 21(2), 72-76.

Geschwind, N., Nicolson, N. A., Peeters, F., van Os, J., Barge-Schaapveld, D., & Wichers, M. (2011). Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. European Neuropsychopharmacology, 21(3), 241-247.

Husky, M. M., Gindre, C., Mazure, C. M., Brebant, C., Nolen-Hoeksema, S., Sanacora, G., & Swendsen, J. (2010). Computerized ambulatory monitoring in mood disorders: feasibility, compliance, and reactivity. Psychiatry Research, 178(2), 440-442.

ICH Expert Working Group. (1998). Statistical principles for clinical trials E9. Paper presented at the International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use.

Littlewood, D. L., Kyle, S. D., Carter, L.-A., Peters, S., Pratt, D., & Gooding, P. (2019). Short sleep duration and poor sleep quality predict next-day suicidal ideation: an ecological momentary assessment study. Psychological medicine, 49(3), 403-411.

Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. Drug and alcohol dependence, 114(2-3), 153-158.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine, 151(4), 264-269.

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annual review of clinical psychology, 13, 23-47.

Myin-Germeys, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. Current opinion in psychiatry, 29(4), 258-263.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & Van Os, J. (2009). Experience sampling research in psychopathology: opening the black box of daily life. Psychological medicine, 39(9), 1533-1547.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry, 17(2), 123-132.

National Heart Lung and Blood Institute. (2018). Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Retrieved from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nelson, J., Klumparendt, A., Doebler, P., & Ehring, T. (2018). Everyday emotional dynamics in major depression. Emotion.

Ottaviani, C., Shahabi, L., Tarvainen, M., Cook, I., Abrams, M., & Shapiro, D. (2015). Cognitive, behavioral, and autonomic correlates of mind wandering and perseverative cognition in major depression. Frontiers in neuroscience, 8, 433.

Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. Emotion, 6(3), 383.

Peeters, F., Berkhof, J., Rottenberg, J., & Nicolson, N. A. (2010). Ambulatory emotional reactivity to negative daily life events predicts remission from major depressive disorder. Behaviour Research and Therapy, 48(8), 754-760.

Putnam, K. M., & McSweeney, L. B. (2008). Depressive symptoms and baseline prefrontal EEG alpha activity: a study utilizing Ecological Momentary Assessment. Biological Psychology, 77(2), 237-240.

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. Psychological assessment, 31(2), 226.

Schwartz, J., Ordaz, S. J., Kircanski, K., Ho, T. C., Davis, E. G., Camacho, M. C., & Gotlib, I. H. (2019). Resting-state functional connectivity and inflexibility of daily emotions in major depression. Journal of Affective Disorders, 249, 26-34.

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC health services research, 17(1), 88.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. Annu. Rev. Clin. Psychol., 4, 1-32.

Simblett, S., Matcham, F., Siddi, S., Bulgari, V., Barattieri di San Pietro, C., López, J., . . . RADAR-CNS Consortium. (2019). Barriers to and facilitators of engagement With mHealth

technology for remote measurement and management of depression: Qualitative analysis. JMIR mHealth and uHealth, 7(1), e11325.

Skinner, B. F. (1965). Science and human behavior: Simon and Schuster.

Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., De Jonge, P., & Wichers, M. (2017). The impact of treatments for depression on the dynamic network structure of mental states: Two randomized controlled trials. Scientific reports, 7, 46523.

Starr, L. R. (2015). When support seeking backfires: Co-rumination, excessive reassurance seeking, and depressed mood in the daily lives of young adults. Journal of Social and Clinical Psychology, 34(5), 436-457.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. Annals of Behavioral Medicine.

Telford, C., McCarthy-Jones, S., Corcoran, R., & Rowse, G. (2012). Experience sampling methodology studies of depression: the state of the art. Psychological medicine, 42(6), 1119-1129.

Torous, J., Lipschitz, J., Ng, M., & Firth, J. (2019). Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. Journal of Affective Disorders.

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 mental health, 2(1), e8.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. Annual review of clinical psychology, 9, 151-176.

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section.

Vachon, H., Bourbousson, M., Deschamps, T., Doron, J., Bulteau, S., Sauvaget, A., & Thomas-Ollivier, V. (2016). Repeated self-evaluations may involve familiarization: an exploratory study related to Ecological Momentary Assessment designs in patients with major depressive disorder. Psychiatry Research, 245, 99-104.

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and Retention With the Experience Sampling Method Over the Continuum of Severe Mental Disorders: Meta-Analysis and Recommendations. Journal of medical Internet research, 21(12), e14475.

van Genugten, C. R., Schuurmans, J., Lamers, F., Riese, H., Penninx, B. W., Schoevers, R. A., . . Smit, J. H. (2020). Experienced Burden of and Adherence to Smartphone-Based Ecological Momentary Assessment in Persons with Affective Disorders. Journal of clinical medicine, 9(2), 322.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics, 30(3), 261-293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of statistical software, 36(3), 1-48.

Watson, L. A., Berntsen, D., Kuyken, W., & Watkins, E. R. (2012). The characteristics of involuntary and voluntary autobiographical memories in depressed and never depressed individuals. Consciousness and cognition, 21(3), 1382-1392.

Figure 1. Flowchart of the systematic review

Figure 2. Forest plot of dropout rate across studies

Table 1. Studies' data included in the analyses

Authors, Year	**MDD Participants**			**ESM/EMA techniques**					**Benefits**	**Dropout information**		
	*N**	*Males (%), mean age (SD)**	*Depression severity**	*Sampling method*	*Type of device*	*Duration, days (n)*	*Number of assessments each day (n)*	*Number of items each day (n)*	*Monetary incentive*	*Dropout (n)*	*Dropout rate (%)*	*Reasons for dropping (n)*
Barge-Schaapveld et al., 1995	25	Nr	Nr	Time-based	Paper & pencil + Electronic	6	10	Nr	No	4	16%	Adverse events not related to treatment (n=2); difficulties complying with ESM/EMA procedures (n=2)
Ben-Zeev et al., 2009	26	19.2%, 40.5 (12)	Severe	Time-based	Electronic	7	9	90	Yes	3	11.5%	Nr
Cohen et al., 2008	76	Nr	Nr	Time-based	Electronic	7	1	26	No	6	7.9%	Nr
Conrad et al., 2008	48	Nr	Nr	Time-based	Electronic	Na	6	60	Yes	2	4.2%	Technical difficulties related to ESM/EMA (n=2)
Dang et al., 2016	4	Nr	Nr	Event and time-based	Electronic	7	Nr	Nr	No	0	0%	NA
Geschwind et al., 2011	74	Nr	Nr	Time-based	Paper & pencil + Electronic	6	10	160	No	11	14.9%	Nr
Husky et al., 2010	20	25%, 42.2 (10.47)	Nr	Time-based	Electronic	3	5	Nr	No	0	0%	NA
Kim et al., 2015	14	85.7%, 34 (5.7)	Mild	Event and time-based	Electronic	37	10	94	No	0	0%	NA
Littlewood et al., 2018	54	Nr	Nr	Event and time-based	Electronic	7	6	28	Yes	3	5.6%	Equipment failure problems (n=2); anxiety due to the random timing of the sampling schedule (n=1)

Nelson et al., 2018	44	Nr	Nr	Time-based	Electronic	4	10	120	Yes	4	9%	Lack of compliance (n=3); technical problems (n=1)
Ottaviani et al., 2015	18	33.3%, 38.4 (12.1)	Severe	Event and time-based	Electronic	1	31	330	Yes	0	0%	NA
Peeters et al., 2006	47	43.5%, 40 (11)	Moderate	Time-based	Paper & pencil + Electronic	6	10	163	No	1	2.1%	Nr
Putnam et al., 2008	6	33.3%, 32.6 (12.1)	Severe	Time-based	Electronic	7	5	Nr	No	0	0%	NA
Schwartz et al. 2019	33	30.3%, 26.01 (3.71)	Moderate	Time-based	Electronic	21	5	90	Yes	0	0%	NA
Snippe et al., 2017	74	Nr	Nr	Time-based	Electronic	6	11	55	No	11	14.9%	Nr
Starr, 2015	11	Nr	Nr	Time-based	Electronic	14	1	24	Yes	0	0%	NA
Torous et al., 2015	13	23%, 48 (16)	Mild	Time-based	Electronic	30	3	9	Yes	0	0%	NA
Vachon et al., 2016	28	Nr	Nr	Time-based	Electronic	150	2	14	Yes	4	14.3%	Lack of compliance (n=2); voluntary dropout (n=1); adverse personal event (n=1)
Watson et al., 2012	24	Nr	Nr	Time-based	Paper & pencil	14-30	Nr	Nr	Yes	4	16.7%	Researcher time constraints (n=1); nr (n=3)

*referring to the participants with a diagnosis of MDD which started the EMA/ESM monitoring.

NA= Not applicable; Nr= Not reported.

Table 2. Results of the subgroup analysis

Study subgroups	N. of studies	Dropout rate		Heterogeneity		
		% [95% CI]	p-value	Group heterogeneity*		
				Q	df(Q)	p-value
Total	19	3.64 [1.37; 5.90]	**0.002**	60.6	18	<0.001
Level of depression severity assessment						
Mild	2	0.00 [-0.16; 0.16]	0.982	0.0	1	0.999
Moderate	2	0.02 [-0.40; 0.45]	0.915	1.0	1	0.312
Severe	3	0.01 [-0.32; 0.34]	0.950	3.4	2	0.184
Overall (residual) heterogeneity				4.4	4	0.353
Sampling method						
Time-based	15	5.22 [2.08; 8.36]	**0.001**	57.4	14	<0.001
Event and time-based	4	0.02 [-0.29; 0.33]	0.917	3.2	3	0.366
Overall (residual) heterogeneity				60.6	17	<0.001
Type of devices						
Paper and pencil forms + Electronic device	3	9.71 [0.07; 19.35]	**0.048**	9.7	2	0.008
Electronic devices only	15	0.04 [0.16; 0.25]	0.673	37.1	14	0.001
Overall (residual) heterogeneity				46.9	16	<0.001
Monetary incentive						
Yes	10	2.22 [0.01; 4.44]	**0.049**	22.5	9	0.007

No	9	4.52 [0.48; 8.56]	**0.028**	38.1	8	<0.001
Overall (residual) heterogeneity				60.6	17	<0.001

Table 3. Results of the meta-regression analysis (univariate)

Risk Factor	Type of data	Number of studies	Estimate	p-value	Explained heterogeneity R^2 (%)	Test for residual heterogeneity		
						Q_E	df(Q)	p-value
Age	Mean age	8	0.0000	0.985	0.0	4.4	6	0.621
Sex	% of males	8	0.0000	0.994	0.0	4.4	6	0.621
Duration of the assessment	N of days	16	0.0003	0.511	0.0	50.6	14	<0.001
Number of samplings	N of samplings each day	17	-0.0002	0.899	0.0	55.8	15	<0.001
Number of items	N of items each day	14	-0.0001	0.598	0.0	51.0	12	<0.001

Table 4. Methodological quality assessment of the included studies

Authors, Year	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	# Yes	Rating
Barge-Schapveld et al., 1995	Yes	No	Yes	UN	No	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	No	9	good
Ben-Zeev et al., 2009	Yes	No	No	UN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	Yes	11	good
Cohen et al., 2008	Yes	No	Yes	UN	No	Yes	Yes	No	Yes	Yes	Yes	UN	Yes	Yes	9	fair
Conrad et al., 2008	Yes	No	No	UN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	Yes	10	good
Dang et al., 2016	Yes	No	UN	Yes	No	Yes	UN	UN	Yes	Yes	Yes	No	Yes	No	7	fair
Geschwind et al., 2011	Yes	Yes	UN	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	No	9	fair
Husky et al., 2010	Yes	No	UN	UN	No	Yes	Yes	Yes	Yes	Yes	Yes	UN	No	Yes	8	fair
Kim et al., 2015	Yes	No	UN	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	9	fair
Littlewood et al., 2018	Yes	Yes	UN	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	Yes	11	good
Nelson et al., 2018	Yes	Yes	UN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	UN	Yes	Yes	11	good
Ottaviani et al., 2015	Yes	No	UN	UN	No	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	7	fair
Peeters et al., 2006	Yes	No	Yes	UN	No	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	Yes	10	good
Putnam et al., 2008	Yes	No	UN	UN	No	Yes	Yes	Yes	Yes	Yes	Yes	UN	Yes	Yes	10	good
Schwartz et al. 2019	Yes	No	UN	Yes	No	Yes	Yes	Yes	Yes	Yes	No	UN	Yes	Yes	9	fair
Starr, 2015	Yes	No	UN	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	10	good
Snippe et al., 2017	Yes	Yes	Yes	Yes	No	UN	UN	No	No	Yes	Yes	UN	Yes	No	7	fair
Torous et al., 2015	Yes	No	Yes	UN	No	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	8	fair
Vachon et al., 2016	Yes	No	Yes	Yes	No	No	Yes	No	No	Yes	No	UN	Yes	No	6	poor
Watson et al., 2012	Yes	Yes	No	UN	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	9	fair

Note. UN (cannot determine/not reported/not applicable). Q=Question of the NIH Quality Assessment Tool for Observational Cohort ad Cross-Sectional Studies; Q1 = research question or objective clearly stated; Q2 = study population clearly specified and defined; Q3 = participation rate at least 50%; Q4 = participants recruitment from similar population; Q5 = sample size justifications provided; Q6 = exposure(s) of interest measured prior to outcome assessment; Q7 = timeframe sufficient; Q8 = different levels of exposures considered; Q9 = exposure measures consistently implemented; Q10 = multiple measurements of exposure; Q11 = outcome measures clearly defined; Q12 = blindness of assessors; Q13 = loss to follow-up 20% or less; Q14 = potential confounding factors included.