Methodological strategies for ecological momentary assessment using mobile phones to evaluate mood and stress in subclinical adult populations: Integrative review

Abstract

**Background:** Ecological momentary assessment (EMA) has utility for measuring psychological properties in daily life. EMA has also allowed researchers to collect data on diverse experiences and symptoms from various subjects.

**Objective:** To review the methodological strategies and usefulness of EMA using mobile phone applications to assess mood and stress in subclinical subjects.

**Methods:** We searched PubMed, CINAHL, EMBASE, the Cochrane Library, PsycINFO, and Web of Science. This review included studies published in peer-reviewed journals in English between 2008 and November 2017 that used mobile phone applications to measure momentary mood or stress in adults. We excluded studies using ecological momentary interventions, studies of smoking, addictions, major psychiatric disorders, and studies measuring non-clinical variables for non-clinical subjects.

**Results:** We reviewed nine selected articles that used EMA via mobile applications to measure momentary mood and stress, and other related variables from various subjects, such as people with chronic fatigue syndrome, breast cancer, migraine, HIV, depressed mood during pregnancy, and adults with mood changes. Most of the selected studies used signal-contingent at random or semi-random intervals to prompt the momentary measurement. Seven of nine studies used specific
applications directly installed on mobile phones and the remainder used mobile phones to link to web-based survey programs.

**Conclusions:** The current study provides researchers with useful information regarding methodological details for utilizing EMA to measure mood and stress. This review shows that EMA methods could be effective and reasonable for measuring momentary mood and stress, given that smartphones are ubiquitous in the general population. Therefore, researchers could adopt and utilize EMA methods to measure psychological health outcomes, such as mood and stress, in subclinical populations.

**Keywords:** Review; experience sampling method; ecological momentary assessment; mobile applications; affect; stress
Introduction

Momentary assessment techniques, such as the ecological momentary assessment (EMA) or experience sampling methods (ESM), have a long tradition [1]. Originally, paper diaries were used in combination with pagers or electronic wristwatches. As technology became more advanced, data collection logistics and reliability were improved by the use of personal digital assistants and smartphone applications [2]. The method focuses on symptoms and adaptive function, such as well-being, and aims to map daily psychological function [3]; it has produced many important findings with respect to psychological properties in the daily life of subjects.

EMA has a number of advantages. It has high ecological validity because assessments are made in natural and real-life environments [3, 4], reduces recall bias, and avoids aggregation because it assesses only the actual moment of interest repeatedly at multiple time points. This characteristic increases accuracy [4] and sensitivity to detect changes [5] in psychological properties. Repeated measures over time can reduce assessment error and improve the validity, reliability, and transparency of individual pattern assessments [3].

EMA allows researchers to disentangle and understand the variability in mental states and psychological constructs by providing accumulated data from the repeated assessments in different situations [6]. Furthermore, it yields an abundant data that can include information on mental state, quality of life, mobility, social network and more, which reduces the need to use separate questionnaires that measure different constructs [3].
This method is also considered suitable for understanding daily changes in psychological features such as mood and stress [7, 8]. In particular, EMA has facilitated the real-time assessment of mood and stress by collecting data throughout the day on multiple occasions and in various environments.

Traditionally, mood and stress have been assessed using retrospective measures [9]; these traditional methods may include a risk of recall bias and reduced accuracy [10]. Therefore, EMA methods might provide health care providers with more accurate data. This may increase access to effective treatments by enabling better understanding of the daily mood and stress of subjects, which are closely related to environmental factors.

Recently, EMA has also enabled researchers to observe the momentary experience of stress and affect. There have been systematic reviews of EMA monitoring mood and/or stress for patients with psychological problems such as mood disorders [11], affective disorders [12], anxiety disorders [13]. However, there is no extant review of EMA methodology and its usefulness for measuring stress or affect in subclinical populations. Accordingly, this review paper provides methodological details of EMA technology in assessing mood and stress in subclinical populations beyond psychiatry field.

**Methods**

**Information source and search strategy**

The search included studies that used mobile applications to measure momentary mood or stress in adults, which were published in peer-reviewed
journals in English between January 2008 and November 2017. We searched six online biomedical databases — PubMed, CINAHL, EMBASE, the Cochrane Library, PsycINFO, and Web of Science — using the following search term: ("ecological momentary assessment"[MeSH] OR “experience sampling” OR “ecological momentary” OR “event sampling” OR “ambulatory assessment” OR “structured diary method” OR “real-time data capture studies” OR “real-time data capture study” OR “beeper studies” OR “beeper study” OR “intensive longitudinal assessment”) AND ("stress, psychological" [MeSH] OR “affect” [MeSH] OR “mood” OR “emotion” OR “affection” OR “stress”) AND (“mobile applications” [MeSH] OR “smartphone” [MeSH] OR “cell phones” [MeSH] OR “smartphone*” OR “cell phone” OR “cellular phone” OR “mobile app*”). The articles identified were inspected, including their reference lists and in-text citations of relevant articles.

**Study selection**

Studies were included that used mobile phone applications to measure momentary mood or stress in subclinical adult populations, which were published in peer-reviewed journals in English. Specifically, included studies utilized mobile phones or smart phones or cellular phones and included adult participants without severe mental illness such as anxiety and major depression or addictions. An earliest year of publication of 2008 was chosen because the first application downloaded on a mobile device was in 2008 [14].

Studies were excluded if they were exploratory pilot or feasibility studies, intervention studies using ecological momentary intervention methods, or studies of
smoking, diets, addictions, major psychological problems, and non-clinical problems in the general population.

**Screening procedure**

A total of 755 articles were retrieved from the six databases, in which 248 records were duplicated. Following removal of duplicates, 507 articles were screened based on the titles and abstracts, of which 13 full-text articles were retrieved by two researchers (Yang and Ryu). Finally, nine full-text articles were selected by according to the criteria and relevant data were extracted. Figure 1 shows the process of study selection based on the PRISMA guidelines [15].

**Data extraction**

The following information was extracted: study purpose, sample characteristics, main momentary measurement, data analysis method, and methodological details of EMA, such as operating system, contingency, duration of data collection, frequency/day, and alarm interval for each study.
Figure 1. PRISMA flowchart providing an overview of the study selection process in the current review.
Results

In total, nine studies met the selection criteria. The following sections summarize how EMA approaches were applied to study populations.

Subject characteristics and main measurements

Various study populations were used: people with chronic fatigue syndrome [16, 17], breast cancer [18], migraine [19], HIV [20], depressed pregnant women [21], and subjects with stress or mood problems [22, 23, 24] (Table 1). Eight out of 9 studies [16-23] measured mood or affect: depressed mood in four studies [21-24], and stress or stressors in four studies [19, 20, 23, 24]. Four studies measured physical activity along with mood or stress [16, 17, 21, 24].

One study of patients with chronic fatigue syndrome assessed patient activity management strategies, patient affect, and symptoms to investigate whether activity patterns occurred according to patient symptom experience and affect [16]. The other study examined the relationship between significant others’ responses and patient outcomes such as affect, symptom severity, disability, and activity management strategies [17].

A study of patients with breast cancer measured sleep satisfaction, mood, and anxiety to evaluate the potential of a mobile mental-health tracker using daily mental-health ratings as indicators of depression. The study examined the impact of adherence to reporting using a mobile mental-health tracker on accuracy of depression screening [18].
One study assessed migraine attacks and prodromal features such as fatigue, cognitive functioning, affect, effort spent (e.g., working hard, felt strained, etc.), and stressors, to test and identify individual prodromal features related to the interictal state in adults with migraine [19]. Control beliefs, mood, stress, coping, and social support were assessed in patients living with HIV to examine whether momentary motivation is a mechanism by which everyday experiences affect adherence to medication therapy [20]. Faherty et al. [21] measured mood, time on foot per day (mobility), and farthest distance traveled per day to examine the association between self-reported daily mood and movement patterns of pregnant women at risk for perinatal depression.

In a study of adults vulnerable to depression, sad mood and self-esteem were measured to understand whether mood-reactive self-esteem represents vulnerability to depression [22]. Among 81 participants, 10 were taking psychotropic medication and 26 participants reported symptoms consistent with clinical depression at the baseline assessment. Connolly et al. [23] assessed momentary life stress, rumination, and depressed mood among undergraduate and graduate students experiencing stress to determine the moderation effect of state rumination on stress in predicting depressive symptoms. Fifty of the participants (41%) reported experiencing at least one major depressive episode in their life and 21 subjects (17%) had a history of multiple major depressive episodes.

In a study that tested the relative contributions and time course of proposed risk/protective factors for depressed mood in daily life, EMA was used to measure stress, coping, social interactions, and physical activity EMA [24]. Of the 73
participants, 16 (21.9%) identified that they had previously been diagnosed with
depression, and of the 16 participants, 7 (43.8%) reported that they were currently
receiving treatment for depressive symptoms.

Table 1. Study details including study purpose, sample characteristics, and main
measurements

<table>
<thead>
<tr>
<th>Author(year), country</th>
<th>Study purpose</th>
<th>Sample characteristics and size (Mean age ± SD)</th>
<th>Main momentary measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band et al. (2017), UK</td>
<td>To investigate whether activity patterns occurred according to patient symptom experience and affect</td>
<td>Chronic fatigue syndrome patients n=23 (M=35.50 ± 13.96)</td>
<td>Patient activity management strategies, patient affects, symptoms</td>
</tr>
<tr>
<td>Band et al. (2016), UK</td>
<td>To examine relationship between significant other responses and patient outcomes</td>
<td>Pairs of chronic fatigue syndrome patients &amp; significant others n=23 (M=35.5 ± 13.96)</td>
<td>Affects, significant others’ responses, symptom severity, disability, activity management strategies</td>
</tr>
<tr>
<td>Clasen et al. (2015), USA</td>
<td>To understand whether mood-reactive self-esteem represents vulnerability for depression</td>
<td>Adults vulnerable to depression n=81 (M=28.72 ± 8.15)</td>
<td>Sad mood and self-esteem</td>
</tr>
<tr>
<td>Connolly et al. (2017), USA</td>
<td>To demonstrate that state rumination moderates the effect of stress in predicting depressive symptoms</td>
<td>Undergraduates/graduates n=121 (M=21.7 ± 5.21)</td>
<td>Life stress, rumination, depressed mood</td>
</tr>
<tr>
<td>Cook et al. (2017), USA</td>
<td>To test whether momentary motivation was a mechanism by which other everyday experiences affect adherence</td>
<td>Patients living with HIV n=87 (M=40.0 ± 8.84)</td>
<td>Control beliefs, mood, stress, coping, social support</td>
</tr>
<tr>
<td>Faherty et al. (2017), USA</td>
<td>To examine the association of daily mood and movement patterns of pregnant women at risk for</td>
<td>Depression women in pregnancy n=36 (M=25.7 ± 5.9)</td>
<td>Mood, movement on foot per day[mobility], farthest radius traveled per</td>
</tr>
</tbody>
</table>
perinatal depression
day\[radius\]

<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Participants</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuller-Tyszkiewicz et al. (2017), Australia</td>
<td>To test relative contributions and time course of proposed risk/protective factors influencing depressed mood in daily life</td>
<td>Adults n=73 (M=29.52 ± 7.97)</td>
<td>Tiredness, depressed mood, stress, coping, social interaction, physical activity</td>
</tr>
<tr>
<td>Houtveen et al. (2013), Netherlands</td>
<td>To test prodromal functioning relative to the interictal state in migraine patients</td>
<td>Adults with migraine n=87 (M=44.5) [25-68]</td>
<td>Migraine attacks, prodromal features; fatigue, cognitive functioning, affects, stressors</td>
</tr>
<tr>
<td>Kim et al. (2016), South Korea</td>
<td>To evaluate the potential of a mobile mental-health tracker using daily mental-health rating as indicators for depression, the impact of adherence on reporting and its accuracy in depression screening</td>
<td>Breast cancer patients n=78 (M=44.35 ± 7.01)</td>
<td>Sleep satisfaction, mood, anxiety</td>
</tr>
</tbody>
</table>

**Methodological details of EMA**

Table 2 shows the methodological details of EMA used in the studies, such as the operating system, application, contingency, duration of data collection, frequency/day, and alarm interval. Although different operating systems were used to install the mobile application, more than half of the studies used Android [16, 17, 20-22]. Seven of nine studies used specific applications directly installed onto mobile phones [16-21, 24], while the remaining two used a web-based survey program that was accessed from the mobile phones [22, 23].
Most studies used commercial applications or survey questionnaires [16, 17, 19-24] and one study customized an application developed in previous studies [18]. Only one study developed a new application for its own study [22].

The duration of the study ranged from a minimum of 6 days [16, 17] to a maximum of 70 days [20], and the contingency frequency varied from once per two weeks [18] to ten times per day [16, 17]. Duration and frequency were related to the study duration, whereby shorter studies collected data more frequently.

All of the methodological approaches [16-17, 19-24] used signal contingency to prompt the momentary measurement, except one study that did not specify the method [18]. The interval of the reminder signal varied according to the study design, since it was stratified for semi-random intervals.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Operating system</th>
<th>Application name</th>
<th>Continuity name</th>
<th>Duration</th>
<th>Frequency/day</th>
<th>Total</th>
<th>Interval (time period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band et al. (2017)</td>
<td>Android OS</td>
<td>Clintouch Signal</td>
<td></td>
<td>6 days</td>
<td>10</td>
<td>60</td>
<td>90 mins semi-random (07:30-22:30)</td>
</tr>
<tr>
<td>Band et al. (2016)</td>
<td>Android OS</td>
<td>Clintouch Signal</td>
<td></td>
<td>6 days</td>
<td>10</td>
<td>60</td>
<td>90 mins semi-random (07:30-22:30)</td>
</tr>
<tr>
<td>Clasen et al. (2015)</td>
<td>Android OS &amp; iOS</td>
<td>Linked to online survey</td>
<td>Signal</td>
<td>21 days</td>
<td>5</td>
<td>105</td>
<td>Pseudo-random (waking hours)</td>
</tr>
<tr>
<td>Connolly et al. (2017)</td>
<td>Not specified</td>
<td>Linked to online survey</td>
<td>Signal</td>
<td>7 days</td>
<td>4</td>
<td>28</td>
<td>Minimum of 90 mins random (11:00-23:00)</td>
</tr>
<tr>
<td>Study</td>
<td>Platform</td>
<td>App/Device</td>
<td>Signal</td>
<td>Duration</td>
<td>Frequency</td>
<td>Response Time</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Cook et al. (2017)</td>
<td>Android OS</td>
<td>Apptiv</td>
<td>Signal</td>
<td>70 days</td>
<td>1</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Faherty et al. (2017)</td>
<td>Android OS</td>
<td>Ginger.io</td>
<td>Signal</td>
<td>56 days</td>
<td>1</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Fuller-Tyszkiewicz et al. (2017)</td>
<td>iOS</td>
<td>MoodQ</td>
<td>Signal</td>
<td>7 days</td>
<td>6</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Houtveen et al. (2013)</td>
<td>Nokia PalmOne Treo</td>
<td>ODA</td>
<td>Signal</td>
<td>28 days</td>
<td>4</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2016)</td>
<td>Not specified</td>
<td>Pit-a-Pat</td>
<td>Not specified</td>
<td>48 weeks</td>
<td>1</td>
<td>336</td>
<td></td>
</tr>
</tbody>
</table>

Study completion rates varied from 50% [24] to 89.5% [19] (Table 3). Fuller-Tyszkiewicz et al. provided a gift voucher; their completion rate was 50% [24]. In another study, student participants chose to receive cash or course credit as compensation for participation; the completion rate reached 82% [23]. In one study, incentives based on the completion rate were provided to encourage the participants while EMA was ongoing, but the average completion rate (74%) [22] was not markedly better than in the other studies. Completion rates in other studies that did not mention incentives ranged from 64.6% [17] to 89.5% [19]. Because data sets from EMA studies include diverse sources of variance, such as 1) across individuals, 2) across days, and 3) within days, various analysis methods have been employed to address this complexity and hierarchy of the data. Hence, the majority of the studies used multilevel models (Table 2).
Five of the nine studies reviewed here undertook multilevel modeling analysis [16, 17, 19, 20, 24]. Two studies used the MTMIXED command in Stata for continuous outcome variables in multilevel modeling [16, 17]. Linear mixed-model multilevel analysis with ML estimation was employed in Houtveen and Sorbi [19]. Cook et al. [20] analyzed the relationship between daily variables and motivation using a multilevel modeling framework. One study incorporated process speed analysis, using a three-level multilevel model to determine whether time lag effects were present between variables [24].

Kim et al. [18] estimated random-effects logistic regression parameters and thereafter used receiver operating characteristic (ROC) plots to evaluate the screening accuracy of the model. For repeated measures, linear mixed-effects regression models have been utilized to test the association between mood and movement patterns, considering 1) all daily mood measures, 2) change in daily mood from the prior day, and 3) weekly average of daily mood [21].

Clasen et al. [22] employed dynamic factor modeling — a vector autoregressive (VAR) methodology that utilized structural equation modeling — and a time-lagged covariance matrix in order to properly model temporal covariances in individual time series. Two-level hierarchical linear modeling (HLM) analyses were performed by Connolly et al. [23], in which EMA observations (level 1) were nested within individuals (level 2) in order to differentiate between within-person and between-person variance.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Completion rate</th>
<th>Momentary data analysis method</th>
</tr>
</thead>
</table>

Table 3. Completion rate and method used to analyze momentary data
<table>
<thead>
<tr>
<th>Study</th>
<th>Accuracy</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band et al. (2017)</td>
<td>65.0%</td>
<td>Multilevel models</td>
</tr>
<tr>
<td>Band et al. (2016)</td>
<td>64.6%</td>
<td>Multilevel models</td>
</tr>
<tr>
<td>Clasen et al. (2015)</td>
<td>74.0%</td>
<td>Dynamic factor modeling and time-lagged covariance matrix</td>
</tr>
<tr>
<td>Connolly et al. (2017)</td>
<td>82.0%</td>
<td>Hierarchical linear modeling analysis</td>
</tr>
<tr>
<td>Cook et al. (2017)</td>
<td>73.0%</td>
<td>Multilevel modeling analysis</td>
</tr>
<tr>
<td>Faherty et al. (2017)</td>
<td>Not specified</td>
<td>Linear mixed-effects regression models</td>
</tr>
<tr>
<td>Fuller-Tyszkiewicz et al. (2017)</td>
<td>50.0%</td>
<td>Process speed analysis using multilevel model</td>
</tr>
<tr>
<td>Houtveen et al. (2013)</td>
<td>89.5%</td>
<td>Linear mixed model multilevel analysis</td>
</tr>
<tr>
<td>Kim et al. (2016)</td>
<td>Not specified</td>
<td>random-effect model of logistic regression and receiver operating characteristic (ROC)</td>
</tr>
</tbody>
</table>

Most of the selected studies had a briefing or intake session to ensure that participants understand the EMA application before starting the survey. Participants could practice and ask questions regarding the application during the session. Informed consent and non-EMA measures were also obtained during the session. After finishing the EMA phase, patients were debriefed to evaluate their experiences during the study.

**Discussion**

This review focused on identifying whether EMA methods are useful for measuring stress and mood in subclinical populations and elucidating relevant methodological details. The current review showed that EMA methods have been
used to measure stress and/or mood in diverse subclinical populations, with various ways of setting alarm contingencies, frequencies, and intervals according to the length of the survey and the momentary measurements used.

Traditionally, data on mood and stress have been measured using retrospective measures [9]; these methods may produce increased recall bias and reduced accuracy [10]. Accordingly, EMA methods may be more suitable for understanding daily changes in psychological features [6, 7], such as affect and stress. EMA facilitates the real-time assessment of mood and stress by collecting data on multiple occasions throughout the day and has recently yielded valuable results that have permitted a better understanding of mood and stress. The studies in our review measured mood or affect [16-23] and stress or stressors [19, 20, 23, 24] in various subclinical populations, such as patients with chronic fatigue syndrome, breast cancer, migraine, HIV, pregnant women with depression, and subjects with stress or mood problems. However, there are some limitations to data collection and analysis using EMA methods.

These methods are time consuming and demanding [6]. Ideally, assessments are kept as brief as possible, preferably < 1 minute, with a maximum of 2 minutes [3]. Selection bias is also an issue. Not all patients are willing to participate or comply with the protocol. Although there is no agreed gold standard for an acceptable compliance rate in EMA studies, Stone and Shiffman [25] noted that EMA data would not be representative of participants’ daily lives if compliance were lower than 80%. The present review found a 50–89.5% of completion rate, which was contingent on the nature of the participants.
One challenge is the complexity of EMA data. An EMA protocol usually must consider item selection, period, intensity, signaling algorithm, event recording, application type, and data storage. Our review showed that the frequency of data collection varied from one to ten times per day over a time period of 6 days to 48 weeks. Repeatedly answering the same questions in an EMA application can be frustrating for participants [26]. Related to this complexity of data collection, missing data also present a limitation and could be attributable to cultural or age-related variation [21].

Regarding data analysis, EMA studies tend to produce multilevel data sets from multiple participants who answer a set of questions at multiple times. Therefore, standard linear and logistic regression analysis techniques are insufficient for analysis of EMA data sets. The complexity of EMA data analysis could hinder researchers or clinicians in using this method [5]. This should be taken into account when considering this technology-driven approach.

A limitation of the current review is that we did not include studies that applied interventions using smartphones or mobile phones, i.e. ecological momentary interventions (EMI), to improve mood or stress, since the purpose of this review was to provide insight into methodological strategies for EMA to assess mood or stress.

Conclusions

This review provides researchers with useful information regarding methodological details when utilizing EMA to measure mood and stress in subclinical populations. Despite the limitations of the current study, we believe this
review shows that EMA is an effective and reasonable way of measuring momentary mood and stress in an era in which smartphones are ubiquitous in the general population. In particular, individuals who have experienced mood changes or severe stress can benefit from EMA methods by using commercial mobile applications to monitor or track their mood and stress vulnerabilities. The current review supports the use of EMA methods to evaluate mood and stress. Therefore, researchers could utilize such methods to measure psychological health outcomes, such as mood and stress, in subclinical populations.

**Acknowledgements**
This study was supported by Basic Science Research Program through the National Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030706).

**Conflicts of Interest**
None declared.

**Abbreviations**
EMA: ecological momentary assessment.
EMI: ecological momentary intervention.
References


12. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: Where are we and where are we


