A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people

M. Dubad1*, C. Winsper1, C. Meyer1,2, M. Livanou1 and S. Marwaha1,3

1Mental Health and Wellbeing, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
2Warwick Manufacturing Group, University of Warwick, Coventry, UK
3Affective Disorders Service, Caludon Centre, Coventry, UK

Background. Mobile mood-monitoring applications are increasingly used by mental health providers, widely advocated within research, and a potentially effective method to engage young people. However, little is known about their efficacy and usability in young populations.

Method. A systematic review addressing three research questions focused on young people: (1) what are the psychometric properties of mobile mood-monitoring applications; (2) what is their usability; and (3) what are their positive and negative clinical impacts? Findings were synthesised narratively, study quality assessed and compared with evidence from adult studies.

Results. We reviewed 25 articles. Studies on the psychometric properties of mobile mood-monitoring applications were sparse, but indicate questionable to excellent internal consistency, moderate concurrent validity and good usability. Participation rates ranged from 30% to 99% across studies, and appeared to be affected by methodological factors (e.g. payments) and individual characteristics (e.g. IQ score). Mobile mood-monitoring applications are positively perceived by youth, may reduce depressive symptoms by increasing emotional awareness, and could aid in the detection of mental health and substance use problems. There was very limited evidence on potential negative impacts.

Conclusions. Evidence for the use of mood-monitoring applications in youth is promising but limited due to a lack of high-quality studies. Future work should explicate the effects of mobile mood-monitoring applications on effective self-regulation, clinical outcomes across disorders and young people’s engagement with mental health services. Potential negative impacts in this population should also be investigated, as the adult literature suggests that application use could potentially increase negativity and depression symptoms.

Received 2 September 2016; Revised 19 May 2017; Accepted 23 May 2017; First published online 23 June 2017

Key words: Ecological momentary assessment, mental health, mhealth, mobile application, mood, mood monitoring, young people.

Introduction

Mood is an affective dynamic, which naturally varies across time and contexts (Trull et al. 2015). Problems with regulating mood can play a key role in the development and trajectory of a range of psychopathologies (Paris, 2004; Crowell et al. 2009; Marwaha et al. 2015). Traditionally, mood has been assessed with retrospective measures (Trull et al. 2015). This can increase the risk of recall bias subsequently reducing accuracy (Schwartz et al. 1999; Reid et al. 2009). The relatively recent use of ecological momentary assessment (EMA) facilitates the real-time assessment of mood by collecting data on multiple occasions throughout the day (Wenze & Miller, 2010). Thus, it may be more suitable for understanding daily mood changes (Cristobal-Narvaez et al. 2016; Myin-Germeys et al. 2016; van Knippenberg et al. 2016).

Various EMA techniques exist, ranging from paper-and-pencil to physiological assessment (Wenze & Miller, 2010) to digital data collection. A number of UK governmental reports (HM Government, 2011; Department of Health, 2013) highlight the benefits of digital tools and Information and Communications Technology (ICT) in aiding the objective, reliable assessment and care of mental health problems. With demand for mental health services outgrowing available resources (Department of Health, 2013), technology might relieve some of this pressure by providing
remote resources that increase access to effective treatment while reducing clinician load.

Applications (‘apps’) offer great promise to young people who are disproportionately affected by mental illness or may struggle to engage with mental health services (Seko et al. 2014). Apps are delivered in a medium young people are familiar with. Figures from Ofcom (2015) indicate that 90% of youth between the ages of 16 and 24 own a smartphone, regardless of sociodemographic domain. Given this widespread ownership and apparent attachment to mobile technology (Ofcom, 2015), youths might feel more comfortable with assessments and treatments utilising mobile apps.

Mental health services increasingly use apps (Olff, 2015), many of which have the capacity for EMA to monitor mood (e.g. Sandstrom et al. 2016b). Several reviews with mainly adult studies (e.g. Donker et al. 2013; Naslund et al. 2015; Nicholas et al. 2015; Torous & Powell, 2015; Bakker et al. 2016; Faurholt-Jepsen et al. 2016; Walsh et al. 2016) have appraised evidence for the use of mood-monitoring apps.

Studies included in these reviews provide some evidence for the psychometric properties, e.g. internal consistency (Palmier-Claus et al. 2012) and concurrent validity (Faurholt-Jepsen et al. 2014) of these apps. There is also evidence for usability (Bardram et al. 2013). Participation rates are generally high across studies sampling adults, ranging from 65% (Depp et al. 2015) to 88% (Ainsworth et al. 2013), though Depp et al. (2012) reported much higher completion rates for paper and pencil compared with app measures (82.9% vs. 42.1%). Evidence also suggests that apps may help people with mental health problems to monitor triggers (Bardram et al., 2013), that the capacity to convey experience can be therapeutic, and that apps could be a useful tool for improving patient–clinician communication (Palmier-Claus et al. 2013).

Less is known about the use of mental health apps, particularly mood-monitoring apps, in youth (10–24 years). A scoping review by Seko et al. (2014) suggested that mood-monitoring apps are positively perceived by youth (Matthews et al. 2008a), may improve treatment adherence (Matthews et al. 2008b) and possibly improve mental wellbeing (Kauer et al. 2012). While intriguing, findings were preliminary due to the low quality of available evidence (NCCMH, 2014), the small number of studies on mood-monitoring apps specifically and the limited number of apps studied (n=2) (NCCMH, 2014; Seko et al. 2014).

In summary, mood-monitoring apps offer a potentially important step change in the assessment of mood and delivery of youth mental health services. Despite this potential and the widespread advocacy for their use (e.g. Firth et al. 2016; Sandstrom et al. 2016a), there are no extant reviews examining the psychometric properties, usability and clinical impacts of mood-monitoring apps in young populations. Therefore, a systematic review was completed to address the following research questions: (1) what are the psychometric properties of mobile mood-monitoring apps; (2) what is their usability; (3) and what are their positive and negative clinical impacts among clinical and non-clinical youth populations? Our secondary aims were to frame our findings within the adult literature, and conduct a quality assessment to examine potential sources of bias.

Method
Following a scoping review, the authors developed the protocol delineating the planned methodology. The review was conducted in adherence to this protocol, and in line with the PRISMA statement (Moher et al. 2009).

Information sources and search strategy
The following sources were searched: Medline, EMBASE, PsycINFO, ProQuest Dissertations & Theses, ProQuest SciTech Collection, the Association for Computing Machinery (ACM) Guide to Computing Literature and Web of Science for articles published from 2008 [the year when the first app was launched (Donker et al. 2013)]. Search terms were informed by previous reviews (Seko et al. 2014), and modified following advice from a medical librarian and field experts. The search was conducted by combining five groups of terms (see online Supplementary Table S1) relating to: type of technology (e.g. ‘mhealth’), type of assessment (e.g. ‘ambulatory assessment’), mood-related outcome or problem (e.g. ‘bipolar disorder’), youth population (e.g. ‘youth’), usability/treatment-related outcomes and psychometric properties (e.g. ‘reliability’, ‘validity’). We were interested in all forms of validity potentially examined in the app literature, e.g. concurrent, face or predictive (Faurholt-Jepsen et al. 2016), though we anticipated a paucity of studies due to the novelty of the field. We defined the ‘usability’ of mood-monitoring apps in accordance with the International Organisation for Standardisation (2001) definition of usability, i.e. ‘the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions’. Consistent with previous systematic reviews (Donker et al. 2013), we included young people’s participation rates (i.e. compliance, response and completion) and how apps were perceived by youths (including their acceptability – how satisfied they were with the app, whether it could be used with ease) as markers of usability.
MD conducted a hand search of articles published in *Cyberpsychology, Behavior and Social Network*, the *Journal of Medical Internet Research* (JMIR), the *JMIR Mental Health*, and the *JMIR mHealth and uHealth* over the last 5 years. An additional search of the first 15 pages of Google Scholar was conducted (search terms ‘mood’, ‘phone’, ‘app’ and ‘monitoring’). Reference lists and in-text citations of relevant articles were inspected. Finally, subject experts were approached to identify additional articles.

**Study selection**

Inclusion criteria were:

1. Apps must have been developed for, and delivered through, mobile phones or smartphones;
2. Participants aged 10–24 years (consistent with the World Health Organisation’s definition of young people; World Health Organisation, 1986);
3. Studies included published and unpublished research reported in the grey literature;
4. Studies must have been published in the English language;
5. Studies must have been published in 2008 or later;
6. Studies must have included community or clinical populations (to ensure the inclusion of sub-clinical youth, who may subsequently access care).

**Screening procedure**

Following removal of duplicates, MD and ML independently screened 100% of titles and abstracts for full-text retrieval. MD assessed full-text articles against the inclusion criteria and extracted relevant data.

**Quality assessment**

MD evaluated the quality of included studies for potential risk of bias using Cochrane’s risk of bias tool, in which studies are allocated a rating of high, low or unclear risk of bias (Higgins et al. 2011).

**Data synthesis**

Quantitative and qualitative data were synthesised narratively.

**Results**

**Study selection**

A total of 1747 articles were identified in the initial search, and 19 from the hand search (Fig. 1). Following removal of duplicates, 1176 abstracts were screened, 86 of which were selected for full-text retrieval. There was a high level of agreement between raters ($k = 0.90$). In total, 64 articles were excluded following full-text review. Three additional articles were identified following inspection of included studies. Twenty-five articles were included in the final review.

**Study characteristics**

Table 1 outlines study methodology, the characteristics and features assessed in the studies, and main findings. Three studies reported on a randomised controlled trial (RCT): one was the primary RCT (Reid et al. 2011), and two reported secondary analyses with the same dataset (Kauer et al. 2012; Reid et al. 2013). The remaining studies were non-experimental or quasi-experimental. The search identified 19 published studies and six unpublished studies (four conference proceedings; two theses). The majority of studies ($n = 16$) were quantitative; the remaining nine employed mixed methods.

Sample size ranged from 6 to 1 089 963 participants. Eight studies recruited healthy participants. Eleven studies recruited participants from clinical populations including youth with a range of mental health, emotional or behavioural problems, such as depression ($n = 8$), high-functioning autism/Asperger’s disorder ($n = 2$) and substance or alcohol use ($n = 1$). The remaining six studies recruited participants from mixed populations comprising healthy, mentally ill or substance-using individuals. Mean ages across studies ranged from 10.95 to 23.7 years.

Methods across studies varied greatly. For example, some studies lent participants a phone, whereas others let participants use their own device. Please see Table 1 for a description of the different data collection methods used in each study. As observed in the adult literature, terminology also varied greatly across studies (please see Usability section for more details).

Various apps were used, the most frequent of which was the ‘Mobiletype’ programme (Reid et al. 2009). Mood outcomes were either direct mood assessments, or described mood-related constructs or behaviours (e.g. stress, hostility). Outcomes were monitored over variable time periods. The shortest period was 24 h (Bossmann et al. 2013), the longest 326 days (Matthews & Doherty, 2011). Monitoring schedules also varied, and could comprise hourly, daily or weekly monitoring, or requirements to complete measures a fixed number of times per day (with or without pre-specified time intervals). Reimbursements or incentives were available in 18 studies (e.g. payments, gift vouchers).

**Psychometric properties of mood-monitoring apps**

Nine studies reported on the reliability or validity of mood-monitoring apps.
Reliability

The internal consistency (correlation between items within a scale) was assessed in four studies (Dunton et al. 2011, 2014; Huh et al. 2014; Ansell et al. 2015). As demonstrated in Table 2, levels ranged from questionable to excellent (George & Mallery, 2003).

Validity

Concurrent validity. Three studies examined concurrent validity (the correlation between an assessment and a previously validated assessment of the same construct). Concurrent validity was mostly moderate across studies (see Table 1). Khor et al. (2014a) compared relationships between participant and parent-reported data from the retrospective Responses to Stress Questionnaire (Connor-Smith et al. 2000) and mobile app data recording participants’ responses to stress. In two studies of university students, Ben-Zeev et al. (2015) and Wang et al. (2014) compared momentary app and retrospective questionnaire data on perceived stress.

Face validity. Two studies described participants’ views on the face validity of the ‘Mobiletype’ app (see Table 1 for numerical details). Reid et al. (2012), using a sample with various mental health problems, found that the app was relatively successful in capturing participants’ feelings and current situation. Khor et al. (2014a), using.

Fig. 1. Flowchart of literature search results and selection of studies.
Table 1. Study details including the author (year,) study purpose, sample characteristics, intervention details and a summary of the main findings

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Study purpose</th>
<th>Sample characteristics</th>
<th>Intervention*</th>
<th>Main findings</th>
</tr>
</thead>
</table>
| Ansell et al. (2015) | To explore the effects of marijuana use on impulsivity and hostility in everyday life using smartphone-based EMA | • Sample size: N = 43 (M = 23.7 years)  
• Population type: young recreational substance users  
• Comparison/control: none  
• Location: USA  
• Data collection: in-person research | • App name: not specified  
• Operation system: not specified  
• Accessibility: no web/general/app store access  
• Device: not specified  
• Measurements: daily alcohol, tobacco and marijuana use; daily impulsivity and daily interpersonal hostility  
• Monitoring period: 14 days, monitoring schedule varied. Compliance monitored for irregularities by research staff  
• Incentive/reimbursement: payments + bonus payment for 95% survey responderate | Psychometric properties:  
• Reliability: acceptable to excellent internal consistency  
Usability:  
• Participation rate: impulsivity: 96% completed data; interpersonal interactions: >99% completed data  
Clinical impacts:  
• Potential implications for problems with (perceived) interpersonal hostility |
| Bachmann et al. (2015) | To examine the usability and unobtrusiveness of the Mobile Ambulatory Mood Assessment (MoA²) app | • Sample size: N = 9 (M = 23.4 years)  
• Population type: healthy/ non-clinical participants  
• Comparison/control: none  
• Location: Germany  
• Data collection: in-person research | • App name: MoA²  
• Operation system: Android  
• Accessibility: no web/general/app store access  
• Device: participants used study phones (Google Nexus 4) or personal Android smartphone  
• Measurements: mood, tiredness and stress level  
• Monitoring period: 12 prompts p/day for 4 days  
• Incentive/reimbursement: no payment | Psychometric properties: not studied  
Usability:  
• Participants’ perception: app perceived as user-friendly and convenient  
Clinical impacts: not studied/reported |
| Ben-Zeev et al. (2015) | To examine if smartphone sensor data can be used to measure behaviour and mental health | • Sample size: N = 47 (M = 22.5 years)  
• Population type: students reporting varying levels of depression symptoms  
• Comparison/control: none  
• Location: USA  
• Data collection: in-person research | • App name: StudentLife  
• Operation system: Android  
• Accessibility: no web/general/app store access  
• Device: participants were offered an Android study smartphone – type not specified  
• Measurements: momentary stress and automated sensor data  
• Monitoring period: 10 weeks (sensor data gathered automatically; stress ratings completed daily, 5 days per week)  
• Incentive/reimbursement: (Raffle) prizes | Psychometric properties:  
• Concurrent validity: significant moderate relationship between averaged app-assessed stress ratings and retrospective post-study questionnaire scores on a measure of perceived stress (r = 0.41, p < 0.01)  
Usability:  
• Participation rate: average weekly response rate was 4.92 days a week (98.4%)  
Clinical impacts: not studied/reported |
Mood-monitoring apps: psychometric properties, usability and clinical impacts

**Mood-monitoring apps: psychometric properties, usability and clinical impacts**

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience movisens Edition version 594

**Sample size:** N = 21.4

**Population type:** healthy/ non-clinical students

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience

**Sample size:** N = 119 (M = 10.95 years)

**Population type:** children with varying body mass index (BMI) levels

**Operation system:** Windows

**Accessibility:** web/general access only

**Device:** participants were lent an HTC Shadow.

**Psychometric properties:**

- **Reliability:** acceptable to good internal consistency

---

**App name:** MyExperience movisens Edition

**Sample size:** N = 41 (M = 15.4 years)

**Population type:** young people with varying levels of alcohol intake

**Location:** Australia

**Data collection:** in-person research

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** Addiction Comprehensive Health Enhancement Support System (ACHESS)

**Sample size:** N = 29 (M = 16.6 years)

**Population type:** adolescents with different clinical problems

**Comparison/control:** none

**Location:** USA

**Data collection:** in-person research

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - 89% of assessments completed

**Clinical impacts:**

- Potential implications for relapse prevention

---

**App name:** MyExperience

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** Mobiletype

**Sample size:** N = 119 (M = 10.95 years)

**Population type:** children with varying body mass index (BMI) levels

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were lent a Nokia 6630

**Measurements:** activities, company, location, mood, responses to stressful events and coping, and questions on participants’ previous evening’s alcohol and cannabis use

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

---

**App name:** Health Enhancement Support System

**Sample size:** N = 41 (M = 15.4 years)

**Population type:** young people with varying levels of alcohol intake

**Comparison/control:** none

**Location:** Australia

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** no web/general/app store access

**Device:** participants were lent a Nokia 6630

**Measurements:** activities, company, location, mood, responses to stressful events and coping, and questions on participants’ previous evening’s alcohol and cannabis use

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

---

**App name:** MyExperience movisens Edition

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report

---

**App name:** MyExperience

**Sample size:** N = 62 (M = 21.4 years)

**Population type:** healthy/ non-clinical students

**Comparison/control:** none

**Location:** Germany

**Data collection:** in-person research

**Operation system:** Android

**Accessibility:** web/general access only

**Device:** participants were provided with an HTC Touch 2 smartphone

**Measurements:** valence, calmness and energetic arousal

**Psychometric properties:** not studied

**Usability:**

- **Participation rate:**
  - Mean completion rate was 10.5 electronic diaries per participant
  - Please note that 15 participants were excluded for missing data

**Clinical impacts:** not studied/report
<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Study purpose</th>
<th>Sample characteristics</th>
<th>Intervention</th>
<th>Main findings</th>
</tr>
</thead>
</table>
| Dunton et al. (2011) | To assess if the level and experience of children’s leisure-time physical activity vary with social and physical contexts by means of EMA | - Sample size: N = 121 (M = 11.02 years)  
- Population type: children with varying BMI levels  
- Comparison/control: none  
- Location: USA  
- Data collection: in-person research | - Measurements: main activity type, social context, physical location, mood and enjoyment  
- Monitoring period: monitoring period: 3–7 random prompts p/day within pre-specified times over two data collection waves (duration: 4 days per wave), separated by 6 months  
- Incentive/reimbursement: up to $40 (compensatory) payment | Usability:  
- Participation rate: 76% of assessments completed on average  
Clinical impacts: not studied/reported |
| Huh et al. (2014) | To examine the contextual antecedents to smoking in a sample of Korean American young adult smokers through EMA | - Sample size: N = 22 (M = 21.23 years)  
- Population type: young adult smokers  
- Comparison/control: none  
- Location: USA  
- Data collection: in-person research | - App name: ActiPal (MEI Ltd.)  
- Operation system: Android  
- Accessibility: web/general access only (demo app)  
- Measurements: affect, perceived stress, cigarette craving, and other contextual and environmental measures  
- Device: Android enabled phones (study phones provided if participants owned iPhones)  
- Monitoring period: random non-smoking signal contingent (5+ p/day for 7 days) + event-contingent prompts over a 7 day period. Compliance closely monitored by research staff  
- Incentive/reimbursement: not reported | Psychometric properties:  
- Reliability: questionable to acceptable internal consistency  
- Usability:  
- Participation rate: 92.4% of assessments completed on average  
- Participants’ perception: it should be noted that one participant withdrew from the study due to technical difficulties with the EMA app  
Clinical impacts: not studied/reported |
Kauer et al. (2012)^
A secondary analysis that investigated the relationships between self-monitoring, emotional self-awareness, and depression through EMA

- **Sample size:** \(N = 69\) (\(M = 18.5\) years)
- **Population type:** young people with mild or more severe mental health/emotional problems
- **Comparison/control:** attention comparison \((n = 49, M = 17.4\) years).
- **Location:** Australia
- **Data collection:** in-person research

See Reid et al. (2011)

Kauer et al. (2009)^
To assess the feasibility and usefulness of a mobile phone-based EMA app to gather information on alcohol use and related behaviours

- **Sample size:** \(N = 18\) [mean ages 15.9 years (females) and 15.8 years (males)] in study 1; \(n = 6\) [mean ages 18.3 years (females) and 19.5 years (males)] in study 2
- **Population type:** healthy/non-clinical students in study 1 and high-risk drinkers in study 2
- **Comparison/control:** none
- **Location:** Australia
- **Data collection:** in-person research

Psychometric properties: not studied

Usability:
- **Participation rate:** completion rates were 52.9% for the intervention group and 59.6% for the comparison group

Clinical impacts: implications for depression symptoms

Kenny et al. (2015)
To assess the feasibility of the CopeSmart app

- **Sample size:** \(N = 43\) (\(M = 16.0\) years)
- **Population type:** healthy/non-clinical adolescents
- **Comparison/control:** none
- **Location:** Ireland
- **Data collection:** in-person research

- **App name:** CopeSmart
- **Operation system:** Android + iOS
- **Accessibility:** no web/general/app store access
- **Device:** app was downloaded on participants’ Android or iOS phones
- **Measurements:** happiness, anger, sadness, stress and worries
- **Monitoring period:** 1 week
- **Incentive/reimbursement:** no monetary incentive

Psychometric properties: not studied

Usability:
- **Participation rate:** better compliance for school-based adolescents than older adolescent high-risk drinkers

Clinical impacts: not studied /reported

Khor et al. (2014a)^
To assess the utility of the Mobilitype programme to examine adolescents with High-Functioning Autism/Asperger’s Disorder’s (HFASD) stressors and coping

- **Sample size:** \(N = 31\) (\(M = 14.46\) years) + parents
- **Population type:** adolescents with HFASD
- **Comparison/control:** none
- **Location:** Australia
- **Data collection:** in-person research

- **App name:** Mobilitype (adapted)
- **Operation system:** not specified
- **Accessibility:** no web/general/app store access
- **Device:** participants were lent a Sony Ericson 7501i
- **Measurements:** mood, stress, last time and daily stress
- **Monitoring period:** 4× p/day for 2 weeks

Psychometric properties: not studied

Usability:
- **Participation rate:** participants engaged with the app on 4 out of 7 days (57.1%)
- **Participants’ perception:** the app’s interface layout was liked by 79% of participants. Furthermore, the app was perceived as easy to use (93%); minor technical difficulties with logging on were experienced by 7% of participants; 70% of participants would use the app in the future; 74% believed the app would be used by other young people; and 70% would recommend the app to a friend.

Clinical impacts:
- **Implications for self-awareness**

Khor et al. (2014b)^
To assess the utility of the Mobiletype programme to examine adolescents with High-Functioning Autism/Asperger’s Disorder’s (HFASD) stressors and coping

- **Sample size:** \(N = 70\) (\(M = 15.9\) years) + parents
- **Population type:** adolescents with HFASD
- **Comparison/control:** Health/Emotional Problems (HFASD) Stressors and Coping Scale (HESP) questionnaire
- **Location:** Australia
- **Data collection:** in-person research

- **App name:** Mobiletype
- **Operation system:** not specified
- **Accessibility:** no web/general/app store access
- **Device:** participants were lent a Nokia 6630
- **Measurements:** mood, stress, last time and daily stress
- **Monitoring period:** 4× p/day for 1 week

Psychometric properties: not studied

Usability:
- **Participation rate:** better compliance for school-based adolescents than older adolescent high-risk drinkers

Clinical impacts: not studied /reported

App name: Mobiletype
Operation system: not specified
Accessibility: no web/general/app store access
Device: participants were lent a Nokia 6630
Measurements: mood, stress, last time and daily stress
Monitoring period: 4× p/day for 1 week
Incentive/reimbursement: partial reimbursement/gift voucher (value: $25)
Table 1 (cont.)

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Study purpose</th>
<th>Sample characteristics</th>
<th>Intervention</th>
<th>Main findings</th>
</tr>
</thead>
</table>
| Khor et al. (2014b) | To investigate how daily hassles, coping, and behaviour and emotional problems are related in adolescents with HFASD | See Khor et al. (2014a) | Incentive/reimbursement: partial reimbursement (value: $20) | ○ A significant moderate to strong correlation for the ‘involuntary engagement’ factor: \( r = 0.70, p < 0.01 \); parent report: \( r = 0.48, p < 0.01 \)
○ A significant strong correlation for the ‘primary control engagement coping’ factor: \( r = 0.53, p < 0.05 \)  
Face validity:  
○ The face validity was measured by assessing how well the app captured participants’ current situation, thoughts and feelings
○ The highest ratings were reported for the app’s ability to capture participants’ feelings (67%); followed by its ability to capture participants’ current situation (63%); and finally its ability to measure participants’ thoughts (50%)
Usability:  
○ Participation rate: participants responded to 61.8% of prompts
○ Note that a substantial proportion of participants gradually stopped responding throughout the study; while every participant completed at least one entry on the first day, completion rates reduced to 45% on day 14
○ Also note that there was a significant positive correlation between full scale IQ and compliance rates (\( r = 0.46, p < 0.01 \))
Clinical impacts: not studied/reported
Psychometric properties: not studied
Usability: not studied/reported
Clinical impacts:  
○ Implications for emotional and behavioural problems
Psychometric properties: not studied
Usability:  
○ Participation rate: different compliance rates across app features - no obvious pattern.
Mean normalised compliance for daily registrations of approximately 30%; mean |
<table>
<thead>
<tr>
<th>Loventoft et al. (2012)</th>
<th>To find out whether people treated for depression would be interested in using a smartphone app for support in their daily lives</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size: ( N = 6 ) (ages 17–24, no means reported)</td>
<td>App name: Daybuilder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population type: young people with recent depression treatment</td>
<td>Operation system: Android</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison/control: none</td>
<td>Accessibility: no web/general/app store access</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location: Denmark</td>
<td>Device: participants provided with Android device with app installed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To assess the issues around the use of mobile phones for mood charting with the aim to improve adolescent engagement

Matthews & Doherty (2011)

Sample size: N = 9 (M = 13.78 years)
Population type: young people with depression, mood disorders, self-harm and anger management
Comparison/control: none
Location: Ireland
Data collection: in-person research

App name: Mobile Mood Diary (MMD)
Operation system: not specified
Accessibility: no web/general/app store access
Device: app downloaded on clients' phones
Measurements: energy, sleep and mood + free area for thought entries
Monitoring period: min. 1× p/day for two sessions
Incentive/reimbursement: reimbursement where necessary

Psychometric properties: not studied
Usability:
Participation rate: 65% response on average

Participants' perception:
Participants' perception: participants preferred mobile technology

Clinical impacts:
Implications for treatment

Matthews et al. (2008b)

To explore the effectiveness of mobile phone v. pen-and-paper for mood tracking

Sample size: N = 73 (M = 14.87 years)
Population type: healthy/ non-clinical students
Comparison/control: paper-based diary condition (n = 52)
Location: Ireland
Data collection: in-person research

App name: MMD
Operation system: not specified
Accessibility: no web/general/app store access
Device: app downloaded on students' phones
Measurements: energy, sleep and mood + free area for thought entries
Monitoring period: 1× p/day for 2 weeks
Incentive/reimbursement: none

Psychometric properties: not studied
Usability:
Participation rate: mobile group significantly more responsive than paper-diary group (t = −2.324, p < 0.05)
Participants' perception: participants preferred mobile technology

Clinical impacts: not studied/reported

Reid et al. (2009)†

A study aimed at developing, piloting and reviewing a youth focused mobile phone programme to track young people's experiences in real time

Sample size: focus group (n = 11, mean age not reported) and pilot study [males (n = 5, M = 15.8 years) and females (n = 13, M = 15.9 years)]
Population type: students
Comparison/control: none
Location: Australia
Data collection: in-person research

App name: MDD
Operation system: not specified
Accessibility: no web/general/app store access
Device: app downloaded on students' phones
Measurements: energy, sleep and mood + free area for thought entries
Monitoring period: 1× p/day for 2 weeks
Incentive/reimbursement: none

Psychometric properties: not studied
Usability:
Participation rate: 76% of diaries
However, response rates decreased from 91% on day 1 to 67% on day 7
Of note, one-third of the sample stated that they did not always respond honestly to items if a specific response would result in further questioning

Participants' perception: the study's initial response rate suggested mobile technology may not be preferred or adopted by all young people. Nevertheless, the app was normalised compliance for weekly registrations of approximately 50%
Participants' perception: user experience negatively affected by technological difficulties; clinicians highlighted the usefulness of self-monitoring when combined with therapy.

Clinical impacts:
Implications for treatment

To explore the effectiveness of mobile phone v. pen-and-paper for mood tracking

Sample size: N = 73 (M = 14.87 years)
Population type: healthy/ non-clinical students
Comparison/control: paper-based diary condition (n = 52)
Location: Ireland
Data collection: in-person research

App name: MMD
Operation system: not specified
Accessibility: no web/general/app store access
Device: app downloaded on students' phones
Measurements: energy, sleep and mood + free area for thought entries
Monitoring period: 1× p/day for 2 weeks
Incentive/reimbursement: none

Psychometric properties: not studied
Usability:
Participation rate: mobile group significantly more responsive than paper-diary group (t = −2.324, p < 0.05)
Participants' perception: participants preferred mobile technology

Clinical impacts: not studied/reported

Psychometric properties: not studied
Usability:
Participation rate:
• Participants' completed 76% of diaries
• However, response rates decreased from 91% on day 1 to 67% on day 7
• Of note, one-third of the sample stated that they did not always respond honestly to items if a specific response would result in further questioning
• Participants' perception: the study's initial response rate suggested mobile technology may not be preferred or adopted by all young people. Nevertheless, the app was
<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Study purpose</th>
<th>Sample characteristics</th>
<th>Intervention</th>
<th>Main findings</th>
</tr>
</thead>
</table>
| Reid et al. (2011)* | A randomised controlled trial to investigate some of the mental health benefits of the Mobiletype programme | • Sample size: $N = 68$ ($M = 18.5$ years)  
• Population type: mild/more mental health or emotional problems  
• Comparison/control: comparison programme ($n = 46$, $M = 17.4$ years)  
• Location: Australia  
• Data collection: in-person research | • App name: Mobiletype  
• Operation system: not specified  
• Accessibility: no web/general/app store access  
• Device: participants were lent a Sony Ericsson Z750i mobile phone  
• Measurements: current activities, company, location, mood, recent stressful events, responses to stressful events, alcohol consumption, cannabis use, and sleep, exercise and diet-related questions  
• Monitoring period: min. 2×/day for 2–4 weeks  
• Incentive/reimbursement: partial reimbursement (A$30) and gift cards (A$20) for post-questionnaires completion (maximum A$60) | overall viewed as youth-friendly and non-invasive  
Clinical impacts: not studied/reported  
Psychometric properties: not studied  
Usability:  
• Participation rate: response rates for the intervention group: 52.9%; comparison group: 60.9%  
Clinical impacts:  
• No significant effects on mental health outcomes; potential implications for self-awareness |
| Reid et al. (2013)* | To assess the utility of Mobiletype in a primary care setting (secondary analysis) | See Reid et al. (2011) | See Reid et al. (2011) | | |
| Reid et al. (2012) | To review Mobiletype in clinical settings | • Sample size: $n = 47$ ($M = 15.59$ years)  
• Mental health/clinical status: adolescents with varied (medical) disorders.  
• Comparison/control: none  
• Location: Australia  
• Data collection: in-person research | • App name: Mobiletype  
• Operation system: not specified  
• Accessibility: no web/general/app store access  
• Device: participants were lent a ZTE F85i JAVA MIDP 2.0 phone with $50$ credit  
• Measurements: location, activity, company, mood, stressful events, responses to stressful events, alcohol and cannabis use, sleep, exercise and diet-related questions  
• Monitoring period: four random prompts p/day for 2–4 weeks (min. completion: 1× p/day) | | | |

*The table continues on the next page...
Incentive/reimbursement: none

Usability:

Participation rate: Participants completed 91% of entries in week 1

Clinical impacts:

Potential implications for assessment and management

Psychometric properties: not studied

Usability:

Participation rate: 85–93% response rate across different measures

Participants’ perception:

- App perceived as ‘easy to use’ (95.6%); ‘a little’ to ‘not at all’ irritating (90.3%)
- The monotony of responding to the same survey questions (15%); the high frequency of the pop-up notifications (9%), and the drain on the phone’s battery life (8%) were perceived as irritating. Participants suggested more varied survey questions (23%), fewer crashes, bugs or freezes (9%) and provided suggestions for novel technical features (13%)
- Some participants also enjoyed the user-friendliness of the app (40%) and the pop-up-reminder feature (17%)

Clinical impacts:

Potential implications for self-reflection on emotions or behaviours

Psychometric properties: not studied

Usability:

Participants’ perception: preference for paper-and-pencil tracking by some participants

Clinical impacts: not studied/reported

Psychometric properties: not studied

Tregarthen et al. (2015) To describe a smartphone app for the self-monitoring of eating disorder symptoms, evaluate characteristics of app

Sample size: N = 1 089 996 [M = 22 years (reported by 48 830 users)]

App name: Recovery Record

Operation system: Android + iOS

Incentive/reimbursement: academic credit and/or prize draw

Clinical impacts: not studied/reported
<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Study purpose</th>
<th>Sample characteristics</th>
<th>Intervention$^a$</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>users and assess the feasibility and utilisation of the app for self-monitoring purposes</td>
<td><strong>Population type</strong>: people with varying levels of ED severity</td>
<td><strong>Accessibility</strong>: general/web and app store access</td>
<td>Usability:</td>
<td><strong>Psychometric properties:</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Comparison/control</strong>: none</td>
<td><strong>Device</strong>: own (iOS or Android) smartphone – type not specified</td>
<td><strong>Participation rate</strong>: 89% of participants monitored ≥ 3 meals; 67% continued to monitor at 30 days</td>
<td><strong>Concurrent validity</strong>: significant moderate relationship between averaged app-assessed stress ratings and retrospective post-study questionnaire scores on a measure of perceived stress ($r = 0.41, p &lt; 0.01$)</td>
</tr>
<tr>
<td></td>
<td><strong>Study Location</strong>: USA</td>
<td><strong>Measurements</strong>: meals and eating disorder-related behaviours/cognitions/feelings/urges</td>
<td><strong>Participants' response</strong>: app received high user ratings</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Data collection</strong>: crowd-sourcing</td>
<td><strong>Monitoring period</strong>: overall usage data not available – six monitoring prompts p/day</td>
<td></td>
<td><strong>Participation rate</strong>: response rates for participants who used own phones: 65%; response rates for participants who used study phones: 72%</td>
</tr>
<tr>
<td>Wang et al. (2014)$^c$</td>
<td>To measure university students' mental health, academic performance and behavioural trends using the StudentLife app</td>
<td><strong>Sample size</strong>: $N = 48$ (M = 22.8 years)</td>
<td><strong>Incentive</strong>: none</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Population type</strong>: university students with varying depression scores.</td>
<td><strong>App name</strong>: StudentLife</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Comparison/control</strong>: none</td>
<td><strong>Operation system</strong>: Android</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Location</strong>: USA</td>
<td><strong>Accessibility</strong>: no web/general/app store access</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Data collection</strong>: in-person research</td>
<td><strong>Device</strong>: participants either used their own Android phones (primary users) or were offered an Android Nexus 4a (secondary users)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Measurements</strong>: momentary mood, sleep, social, physical exercise, activity, and behaviour; automated sensor data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Monitoring period</strong>: 10 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Incentive</strong>: (Raffle) prizes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The accessibility of mood-monitoring apps was assessed through a search of Google and three app stores (iTunes, Google Play and Microsoft store) in June 2016.

$^b$ Please refer to Table 2 for coefficient values.

$^c$ These studies utilised the same data.

$^d$ These studies utilised the same data.

$^e$ These studies utilised the same data.

$^f$ These studies partly utilised the same data.

$^g$ These studies utilised the same data.
Participation rates

Twenty-one studies examined participation rates, which ranged from 30% to 99%. Average percentages were not computed in four studies. Instead, these studies described the mean number of diary entries per participant (Bossmann et al. 2013), between-group differences (Matthews et al. 2008; Kauer et al. 2009), or evidence of ongoing compliance (Tregarthen et al. 2015). There was some indication that response rates were higher in studies with incentives. For example, Matthews et al. (2015) offered an incentive of $50 per week, and had a participation rate of 89% (see Table 1 for comparative rates and incentive details). Participation rates also appeared to be affected by response fatigue. In Reid et al. (2009), for instance, response rates decreased from 91% on day 1 to 67% on day 7. Finally, participation rates were potentially affected by sample-specific characteristics. In a study with high-functioning autistic participants, Khor et al. (2014a) found a significant positive correlation between full-scale IQ and compliance rates ($r = 0.46, p < 0.01$).

Participants’ perceptions

Nine studies considered participants’ perceptions of the apps. Three of these studies specifically referred to the ‘acceptability’ of apps. In Dennis et al. (2015), 95% of adolescents felt that the EMA app ‘was not too long’. Tregarthen et al. (2015) measured app utilisation data as a proxy for acceptability. There were over 100 000 users over a 2-year period (with 89% using the application at least three times), which the authors interpreted as a demonstration of broad acceptability. While they did not define acceptability specifically, Reid et al. (2009) concluded that their app was ‘acceptable’ based on the data they captured (e.g. completion rates, participants’ feedback).

Across studies, 93–100% of respondents found apps easy to learn or use (Dennis et al. 2015; Kenny et al. 2015; Sacco, 2015). In addition, participants rated apps as useful (Kenny et al. 2015), convenient, user-friendly (Bachmann et al. 2015), youth-friendly and non-invasive (Reid et al. 2009). Despite these positive experiences, technological difficulties (e.g. software crashes, reduced battery life) were reported to negatively affect user experience and participation (Loventoft et al. 2012; Huh et al. 2014; Dennis et al. 2015; Sacco, 2015). Although most young people reported a preference for mobile phone mood charting in comparison to paper diaries (Matthews et al. 2008b), not all young people preferred mobile technology (Reid et al. 2009; Scotti, 2015). Scotti (2015), e.g. found that several participants from a sub-diagnostic eating disorder sample favoured paper-and-pencil to track their data.

Table 2. Internal consistency coefficients across studies and domains

<table>
<thead>
<tr>
<th>Authors</th>
<th>Positive affect</th>
<th></th>
<th></th>
<th>Negative affect</th>
<th></th>
<th></th>
<th>Perceived stress</th>
<th></th>
<th></th>
<th>Impulsivity</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>WS</td>
<td>BS</td>
<td>O</td>
<td>WS</td>
<td>BS</td>
<td>O</td>
<td>WS</td>
<td>BS</td>
<td></td>
<td>O</td>
<td>WS</td>
</tr>
<tr>
<td>Ansell et al. (2015)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>0.78</td>
<td>0.96</td>
</tr>
<tr>
<td>Dunton et al. (2011)</td>
<td>0.88</td>
<td>–</td>
<td>0.75</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Dunton et al. (2014)</td>
<td>0.87</td>
<td>–</td>
<td>0.74</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Huh et al. (2014)</td>
<td>0.65</td>
<td>–</td>
<td>0.78</td>
<td>–</td>
<td>0.73</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: O, Overall, WS, within-subject level, BS, between-subject level. Internal consistency coefficients values interpretation: ‘$>0.9$ – excellent, $0.8–0.9$ – good, $0.7–0.8$ – acceptable, $0.6–0.7$ – questionable, $0.5–0.6$ – poor and $<0.5$ – unacceptable’ (George & Mallery, 2003, pp. 231).

a sample with high-functioning autism and Asperger’s found that the app was not quite as successful in these domains. In both studies, the apps were less successful in capturing participants’ thoughts.

Usability of mood-monitoring apps

Participation rates

Positive and negative clinical impacts of mood-monitoring apps

Mental health and awareness

Five (two were from the same RCT) studies examined potential clinical impacts of the apps. Reid et al. (2011) found a significant improvement in emotional self-awareness, but no significant improvements in depression, anxiety or stress scores in youth with mental health or emotional problems. In a secondary analysis of the same RCT, Kauer et al. (2012) reported an indirect association between app use and depression.
symptoms via increased emotional self-awareness. The app, however, did not significantly reduce rumination.

Qualitative feedback from two studies also suggested that mood-monitoring apps can help improve self-awareness (Kenny et al. 2015), and self-reflection on emotions or behaviours (Sacco, 2015).

Though they did not test this premise directly, Ansell et al. (2015) hypothesised that app-based monitoring could have promoted self-awareness in participants subsequently reducing (perceived) interpersonal hostility.

In Khor et al. (2014b), parents rated their children with high-functioning autism as showing fewer symptoms of behaviour and emotional problems following use of the self-monitoring app.

**Treatment implications**

Five studies reported results that could have implications for the prevention and treatment of mental health problems. Mobile app data gathered by Dennis et al. (2015) were used to identify high-risk groups for substance use, which could potentially help with relapse prevention. Crooke et al. (2013) suggested that mood-monitoring apps could help investigate adolescents’ motivations for drinking, thus informing the development of interventions.

Qualitative feedback from therapists suggests that the use of mobile apps could help facilitate engagement with participants suffering from various mental health problems (Matthews & Doherty, 2011). Reid et al. (2012) reported that the Mobiletype app facilitated the assessment and management of youth mental health problems and reduced consultation time with paediatricians; the data captured enabled more individually focused consultations, which assisted in rapport building and communication.

In the third of a series of papers detailing their RCT, Reid et al. (2013) explored the potential treatment benefits of ‘Mobiletype’. In comparison to the control programme, the app significantly increased general practitioners’ (GPs) understanding of their patients’ health and current functioning, and aided diagnoses, communication, medication and referrals. However, there was no significant effect on doctor’s confidence, doctor–patient rapport or pathways to care.

Finally, in a conference paper by Loventoft et al. (2012), clinicians highlighted the usefulness of self-monitoring when combined with therapy.

**Quality assessment**

Please see online Supplementary Fig. S1 for an overall depiction of the risk of bias domains across studies.

Risk of selection bias was difficult to assess in many studies, as they often lacked treatment, control or comparison groups. Three studies (all using the same RCT data) were deemed at low risk of selection bias due to a clear description of the randomisation and concealment allocation process (Reid et al. 2011, 2013; Kauer et al. 2012). Two studies were at unclear risk of selection bias because randomised sequence generation and method of allocation concealment were not sufficiently described (Matthews et al. 2009; Reid et al. 2009). One study was considered at high risk of selection bias (Scotti, 2015) as there was no random allocation process for the control condition.

Only the RCT study (three publications) addressed the blinding of participants and personnel, and was thus considered at low risk of performance bias (Reid et al. 2011, 2013; Kauer et al. 2012). The risk of detection bias in these studies was unclear due to a lack of clarity on blinding of outcome assessments.

The risk of attrition bias was difficult to ascertain in three studies. In one study (Kenny et al. 2015), a number of participants were not included in the final sample due to restrictions on school access (no other information was available). Bossmann et al. (2013) excluded 15 participants from the final sample due to ‘missing data’, but did not provide further information, including whether any analyses were performed to address missing data. Reid et al. (2012) was considered at unclear risk of attrition bias, as there was no information on the participants (21%) lost to follow-up. The remaining studies appeared to be at low risk of attrition bias. There was insufficient information to assess the risk of reporting bias in all studies but those of the RCT, which addressed pre-specified outcomes and appeared to be at low risk (Reid et al. 2011, 2013; Kauer et al. 2012). All studies appeared to be at unclear or high risk of other types of bias.

**Discussion**

The aim of this review was to summarise and evaluate evidence for the use of mobile mood-monitoring apps in young people (aged 10–24 years) from clinical and non-clinical populations. We specifically focused on psychometric properties, usability and clinical impacts.

**Psychometric properties of mood-monitoring apps**

Few studies assessed psychometric properties. There was limited evidence for reliability, with four studies demonstrating questionable to excellent levels of internal consistency. Studies examining concurrent \((n=3)\) and face \((n=2)\) validity were also sparse, making it difficult to draw firm conclusions. Face validity findings, e.g. could have been moderated by sample characteristics, e.g. reduced insight in participants with autism (Khor et al. 2014a).
The limited assessment of psychometric properties observed in the youth literature mirrors the adult literature. Evidence for concurrent validity in adult populations is inconclusive (Depp et al. 2012; Palmier-Claus et al. 2012; Faurholt-Jepsen et al. 2014). Inconsistent methodology across these studies, e.g. momentary (Depp et al. 2012) v. retrospective assessments (Faurholt-Jepsen et al. 2014), varying periods between the event and participants’ recollection of the event (Palmier-Claus et al. 2012), likely contribute to variable findings. Previous evidence suggests that real-time mood measurement methods (e.g. EMA) only have a modest correlation with retrospective assessments, such as questionnaires (Ebner-Priemer & Trull, 2009). This leads to the conceptual question of whether retrospective measures are the most appropriate comparators when assessing the validity of mood-monitoring apps. Questionnaires measure an individual’s retrospective view of their mood state over a number of days. While they are subject to recall bias, this bias incorporates other emotional processing (e.g. contexts) that the more instantaneous assessment of mood (e.g. EMA) may not capture, or at least as richly. Thus, the two assessment methods may be measuring different types of affective experience. As it is difficult to draw robust conclusions about the validity of apps using retrospective assessments, future studies should further examine psychometric properties using other sources of comparative data, e.g. active smartphone app data (i.e. app assessments) with passive sensor smartphone data (Nicholas et al. 2015; Sandstrom et al. 2016b), associations with clinical rating scales (Faurholt-Jepsen et al. 2016).

**Usability of mood-monitoring apps**

The usability of mood-monitoring apps was more extensively studied, and overall studies suggest that apps are usable for young people. However, there were some within- and between-study differences in participants’ perceptions of apps, and participation rates.

Generally, participation rates were lower in studies where participants had mental health difficulties (Reid et al. 2011; Kauer et al. 2012), problematic drinking patterns (Kauer et al. 2009) or autism spectrum disorders – especially those with lower IQ (Khor et al. 2014a). In particular, participation levels were low for those living without set routines (Kauer et al. 2009). This is an important consideration, as youths with mood-related problems, e.g. borderline personality disorder, often have disorganised daily routines (Fleischer et al. 2012). This suggests a need to tailor apps for different clinical populations (Kauer et al. 2009).

Some studies indicated that incentives could positively influence participation rates (e.g. Ansell et al. 2015; Dennis et al. 2015). It may not be financially feasible to offer incentives in non-research settings. However, results tentatively suggest that participation rates may be better for mobile apps than traditional paper-based assessments irrespective of incentives (Matthews et al. 2008b). Participation rates for paper-based diaries are as low as 11% (Stone et al. 2003) compared with 30–99% for mood-monitoring apps in the current review. This supports that apps could lead to better adherence rates than non-digital assessment tools in young populations. Factors that could improve participation rates include the use of less intensive assessments (e.g. once-daily rather than multiple times), shorter assessments and the incorporation of staff monitoring or automatic reminders (Hu et al. 2014).

Studies from the adult literature are somewhat congruent in supporting the usability of mood-monitoring apps (Bardram et al. 2013), though evidence suggests that increasing age (e.g. ‘middle age’) may lower likelihood of mood-monitoring app use (Depp et al. 2012). Both adult (Palmier-Claus et al. 2013) and adolescent (Bradford & Rickwood, 2014) populations expressed some reservations about using apps due to the perceived risk of reduced personal contact (Palmier-Claus et al. 2013).

Overall our review demonstrated that young people positively perceive apps (Reid et al. 2009) and would be willing to use this technology in real-life settings (Kenny et al. 2015; Tregarthen et al. 2015). Very few studies considered clinician perspectives on mood-monitoring apps. Matthews & Doherty (2011) found that therapists’ confidence with technology was the biggest barrier to the use of mood apps. More qualitative studies are now needed to further explore young peoples’ (and clinicians’) perceptions (Hollis et al. 2016) to broaden our understanding of factors pertinent to the uptake of mood-monitoring apps in real-life settings.

**Positive and negative clinical impacts of mood-monitoring apps**

Few of the included studies assessed the clinical impacts of the mood-monitoring apps. Although evidence was generally positive (e.g. facilitating assessment, management and GPs’ understanding), most studies relied on subjective participant feedback (Sacco, 2015) rather than RCT methodology with objective outcome measures.

The preliminary evidence (Kauer et al. 2012) very tentatively suggests that electronic mood-monitoring apps could function as an intervention tool (Seko...
et al. 2014; Olff, 2015; Faurholt-Jepsen et al. 2016). Intriguingly, results from the one RCT indicated that mood-monitoring apps might reduce depression in youths by increasing their levels of emotional awareness (Kauer et al. 2012). Similarly, though in a non-experimental study, Khor et al. (2014b) reported that self-monitoring improved parent-reported behavioural and emotional problems in participants with autism. While these results are promising, they require replication and future studies may further explore the mechanisms via which apps could potentially impact on clinical outcomes. One possibility is that mood apps could have a positive impact on clinical symptoms due to patient/participant expectations regarding their benefits. This phenomenon, coined the digital placebo effect, is an overlooked area, which also merits future investigation (Torous & Firth, 2016).

We were unable to fully examine the potential negative impacts of mood-monitoring apps in youth populations, as they were not directly investigated in studies. However, Reid et al. (2009) found that participants did not always respond to questions truthfully to avoid having to answer further questions. Thus, this type of assessment could potentially lead to the inaccurate assessment (and treatment) of mental health problems.

A small number of adult studies report on the negative effects of mood-monitoring apps. There is some suggestion that apps may increase negative reactivity (Ainsworth et al. 2013), increase focus on negative symptoms and thoughts (Palmier-Claus et al. 2013), and potentially maintain depressive symptoms (Faurholt-Jepsen et al. 2015). Given the evidence from the adult literature, research on the possible harmful effects of app use in youths is needed before these tools are routinely used in clinical practice. Part of this endeavour should seek to identify the optimal balance between a monitoring schedule, which accurately captures affective dynamic processes, while minimising respondent workload (Bolger et al. 2003; Trull et al. 2015). This is particularly important, not only because it affects participation rates, but also because the responsibility of self-monitoring could impose a burden on young people (Shiffman et al. 2008), might result in unnecessary pressure (Lupton, 2013; Seko et al. 2014) and exacerbate mental health problems (Conner & Reid, 2012; Faurholt-Jepsen et al. 2015).

Future work may investigate potential ethical issues surrounding the use of mood-monitoring apps. For example, their use could lead to an over-reliance on technology in young populations, which could exacerbate mental health problems (Thomée et al. 2011). There could also be information security-related risks (e.g. digital theft) that could compromise confidentiality (Prentice & Dobson, 2014). Finally, youths could use apps as a replacement for treatment and health monitoring (Tregarthen et al. 2015). Considering the importance of the therapeutic alliance for successful treatment outcomes (Karver et al. 2006), the efficacy of smartphone apps could be reduced if they are used without clinicians’ involvement (Prentice & Dobson, 2014).

Strengths and limitations

As far as we are aware, this is the first review to systematically examine and quality assess the evidence for the psychometric properties, usability and clinical outcomes of mood-monitoring apps in youth. However, our results should be considered through the lens of a number of limitations.

First, despite undertaking a comprehensive search, there were very few high-quality studies available for inclusion in the review. There was only one primary RCT highlighting the need for more trials on the efficacy of mood-monitoring apps in young people. Indeed, our quality assessment indicated that the majority of studies included some form of bias. For example, many studies were at high or unclear risk of sampling (e.g. self-selected samples) and attrition bias. This could have affected the generalisability of our findings or led to an overestimation of positive effects, e.g. our findings may only apply to individuals with less severe psychopathology who are more likely to engage with services.

Second, studies demonstrated a great variability in terminology (especially for implementation outcomes, e.g. acceptability) making interpretations and cross-study comparisons difficult (inconsistent terminology is also a common feature of the adult app literature). For example, we found that ‘acceptability’ was defined very differently across studies, ranging from proxy markers, i.e. utilisation data (Tregarthen et al. 2015) to participants’ experience of burden (Dennis et al. 2015). This highlights the need for more careful delineation and measurement of implementation outcomes in future work (Proctor et al. 2011).

Third, there were large variations in samples and methodologies, again making cross-study comparisons difficult and quantitative synthesis (i.e. meta-analysis) impossible. Thus, some of our conclusions remain tentative pending further rigorous, higher quality research (e.g. RCTs).

Fourth, it should be noted that studies in this review often used apps that were specifically developed for the study, and therefore not publicly available through app platforms (e.g. iTunes). Thus, there is a need for more research to assess the evidence for apps that are freely downloaded and used by youth,
Mood-monitoring apps: psychometric properties, usability and clinical impacts

and whether their use can be incorporated into clinical care (Nicholas et al. 2015).

Clinical and research implications

Mood-monitoring apps could potentially have positive effects in both clinical and sub-clinical youth populations. Indeed, mood-monitoring apps may help youth identify and address burgeoning mental health and substance use problems (Dennis et al. 2015), and possibly utilise more adaptive coping strategies (Kauer et al. 2012). Further research is needed to examine the effects of these apps in samples with serious mental disorders, such as bipolar disorder (Grunerbl et al. 2016), borderline personality disorder (Lederer et al. 2012), substance use problems (Dennis et al. 2015), and psychosis (Ben-Zeev et al. 2014; Palmier-Claus et al. 2014).

Evidence, though limited, suggests that mood-monitoring apps could potentially aid diagnosis and treatment decision-making (Reid et al. 2013). Future studies should explore whether this technology could aid in the assessment of disorders that can be difficult to differentiate [e.g. borderline personality disorder, bipolar disorder (Yen et al. 2015)] by providing rich data about the timing and extent of mood fluctuations.

As technological innovations have been endorsed at a government level, integrating mood-monitoring apps within mental health services may improve access and relieve some of the strain these services are currently experiencing [e.g. by improving access to mental health treatment (Department of Health, 2013)]. However, to date, the potential positive and negative impacts of apps have not been sufficiently investigated in youth.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S0033291717001659.

Acknowledgements

MD is funded by an Economic and Social Research Council (ESRC) Collaborative Award Studentship – ES/J500203/1. Funding for open access publication was kindly provided by Research Councils UK (RCUK).

References


Khor AS, Gray KM, Reid SC, Melvin GA (2014a). Feasibility and validity of ecological momentary assessment in...


how people feel in different locations. *Journal of Research in Personality* [http://dx.doi.org/10.1016/j.jrp.2016.06.004].


