DOI: 10.1111/bdi.13243

REVIEW

BIPOLAR DISORDERS WILEY

Smartphone-based interventions in bipolar disorder: Systematic review and meta-analyses of efficacy. A position paper from the International Society for Bipolar Disorders (ISBD) Big Data Task Force

Gerard Anmella[1] | Maria Faurholt-Jepsen[2] | Diego Hidalgo-Mazzei[1] | Joaquim Radua[3,4,5] | Ives C. Passos[6] | Flavio Kapczinski[7] | Luciano Minuzzi[7] | Martin Alda[8] | Sandra Meier[8] | Tomas Hajek[8,9] | Pedro Ballester[10] | Boris Birmaher[11] | Danella Hafeman[11] | Tina Goldstein[11] | Elisa Brietzke[12] | Anne Duffy[12] | Benno Haarman[13] | Carlos López-Jaramillo[14,15] | Lakshmi N. Yatham[16] | Raymond W. Lam[16] | Erkki Isometsa[17] | Rodrigo Mansur[18] | Roger S. McIntyre[19] | Benson Mwangi[20] | Eduard Vieta[1] | Lars Vedel Kessing[2,21]

[1]Digital Innovation Group, Bipolar and Depressive Disorders Unit, Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain

[2]Copenhagen Affective Disorder research Center (CADIC), Psychiatric Center Copenhagen, Copenhagen, Denmark

[3]Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, IDIBAPS, CIBERSAM, Barcelona, Spain

[4]Early Psychosis: Interventions and Clinical-detection (EPIC) lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[5]Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[6]Laboratory of Molecular Psychiatry and Bipolar Disorder Program, Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento, Centro de Pesquisa Experimental do Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[7]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

[8]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada

[9]National Institute of Mental Health, Klecany, Czech Republic

[10]Neuroscience Graduate Program, McMaster University, Hamilton, Canada

[11]Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[12]Department of Psychiatry, Queen's University, Kingston, ON, Canada

[13]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[14]Research Group in Psychiatry, Department of Psychiatry, Faculty of Medicine, University of Antioquia, Medellín, Colombia

[15]Mood Disorders Program, Hospital Universitario San Vicente Fundación, Medellín, Colombia

[16]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

[17]Department of Psychiatry, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[18]Mood Disorders Psychopharmacology Unit (MDPU), University Health Network, University of Toronto, Toronto, ON, Canada

[19]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[20]Department of Psychiatry and Behavioral Sciences, UT Center of Excellence on Mood Disorders, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

[21]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Gerard Anmella and Maria Faurholt-Jepsen have contributed equally to this work.

Correspondence
Eduard Vieta, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, 170 Villarroel st, 08036 Barcelona, Catalonia, Spain.
Email: evieta@clinic.cat

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; National Council for Scientific and Technological Development; Generalitat de Catalunya; CIBER of Mental Health; Fondo Europeo de Desarrollo Regional; Instituto de Salud Carlos III (ISCIII); Spanish Ministry of Science and Innovation; Spanish Ministry of Health, Grant/Award Number: CM21/00017; Fondo Social Europeo Plus (FSE+)

Abstract

Background: The clinical effects of smartphone-based interventions for bipolar disorder (BD) have yet to be established.

Objectives: To examine the efficacy of smartphone-based interventions in BD and how the included studies reported user-engagement indicators.

Methods: We conducted a systematic search on January 24, 2022, in PubMed, Scopus, Embase, APA PsycINFO, and Web of Science. We used random-effects meta-analysis to calculate the standardized difference (Hedges' g) in pre-post change scores between smartphone intervention and control conditions. The study was pre-registered with PROSPERO (CRD42021226668).

Results: The literature search identified 6034 studies. Thirteen articles fulfilled the selection criteria. We included seven RCTs and performed meta-analyses comparing the pre-post change in depressive and (hypo)manic symptom severity, functioning, quality of life, and perceived stress between smartphone interventions and control conditions. There was significant heterogeneity among studies and no meta-analysis reached statistical significance. Results were also inconclusive regarding affective relapses and psychiatric readmissions. All studies reported positive user-engagement indicators.

Conclusion: We did not find evidence to support that smartphone interventions may reduce the severity of depressive or manic symptoms in BD. The high heterogeneity of studies supports the need for expert consensus to establish ideally how studies should be designed and the use of more sensitive outcomes, such as affective relapses and psychiatric hospitalizations, as well as the quantification of mood instability. The ISBD Big Data Task Force provides preliminary recommendations to reduce the heterogeneity and achieve more valid evidence in the field.

KEYWORDS
bipolar disorder, smartphone interventions, efficacy, engagement, task force

1 | BACKGROUND

Bipolar disorders (BD) are chronic and usually recurrent major mood disorders with onset typically during adolescence and early adulthood, a lifetime prevalence estimated at 2.4%, and a course of disease that entails fluctuations between depressive and manic episodes.[1] Higher recurrences of mood episodes in BD have been related to reduced response and poor adherence to treatments, progressive neuroanatomic brain changes, and cognitive dysfunction; which, in turn, is associated with poor clinical course and functional impairment.[2,3] The recurrent, and sometimes progressive and severe nature of BD, along with high rates of morbidity and mortality,[4] translates into an estimated 8 to 12 years shortened life expectancy,[5] reduced quality of life, and substantial burden of disease, making BD one of the main causes of disability among young people.[6,7]

This recurrent illness course may be stabilized and overall prognosis improved on appropriately selected prophylactic and maintenance pharmacological treatments, and adjunctive psychosocial interventions, which have proven effective in preventing mood recurrences in BD,[8] especially in the early course of the disease.[9,10] However, people with BD often lack insight about their symptoms and the need of treatment, especially in manic phases,[11,12] and access to evidence-based psychosocial interventions for BD remains limited. In addition, common psychiatric clinical monitoring through routine medical visits mainly consists of periodic cross-sectional symptoms assessments that rely on self-reports, posing several limitations due to recall and confirmation bias and misinterpretations.[13] Hence, more effective strategies for the clinical management of BD are imperative.

Advancements in digital technologies might hold potential solutions to the above challenges. Nowadays, smartphones and wearables can capture behavioral, cognitive, and mood information in an objective, continuous, passive, unobtrusive way,[14] which is known as digital phenotyping.[15] In the last 5 years, the use of smartphones and wearables in psychiatric research has become widespread and yielded promising results.[16,17] BD represents the ideal diagnostic framework for digital phenotyping, as its biphasic nature overtly

translates into altered emotion, speech, and behavior.[18] For instance, BD patients usually show overactivity, euphoria, racing thoughts, and increased self-esteem—during manic episodes—in contrast to low energy, depressed mood, inability to concentrate, and feelings of worthlessness—during depressive episodes.[19]

The foregoing biphasic alterations in motor, social and speech activity can accurately reflect the mood state and their digital quantification using smartphones has been precisely correlated to different mood states.[20–22] Smartphones offer unique capabilities for real-time monitoring of depressive and manic symptoms through self-reported, as well as passively collected data—such as speech and activity—to detect and predict phase changes in BD rather than self-assessment data.[23,24] Smartphones may even have the potential to provide early detection of prodromal symptoms between outpatient visits in BD.[25,26]

In addition to providing the afferent limb of assessment (through self-report or automatically collected data), smartphones have the ability to provide the efferent limb of intervention, which can, in turn, be optimally titrated through continuous feedback from digital phenotyping.[27] Recent studies have shown that people with BD are generally predisposed toward smartphone-based interventions.[28] Nevertheless, despite the huge number of smartphone interventions for BD already available in the Google Play and iOS stores, very few meet evidence-based medicine quality standards.[29,30] Smartphone interventions have already shown positive mild-to-moderate efficacy for treating depression,[31] anxiety,[32] and effectiveness in treating post-traumatic stress disorder (PTSD) symptoms.[33] However, the clinical effects of smartphone interventions on BD have yet to be established.

Thus, our main aim was to examine the efficacy of smartphone-based interventions in BD –including preventing mood episodes, reducing psychiatric admissions, (hypo)manic and/or depressive symptoms, and perceived stress, and/or increasing functioning and quality of life, both in randomized clinical trials (RCT) and observational studies.

As secondary aims, we sought to explore which aspects of smartphone interventions were associated with greater or lesser efficacy and effectiveness, to which subpopulations of people with BD, and how the included studies reported and measured user-engagement indicators (UEIs) for smartphone interventions.

2 | METHODS

This systematic review and meta-analyses followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement for transparent and comprehensive reporting of methodology and results[34] (Table S1). The study protocol was pre-registered with PROSPERO (CRD42021226668).

2.1 | Search strategy

We conducted a comprehensive systematic literature search to examine the efficacy and effectiveness of smartphone-based interventions in BD. The first search was conducted on January 21, 2021 and later updated on January 24, 2022. The search had no year or language restrictions and included the following databases: PubMed, Scopus, Embase, APA PsycINFO, and Web of Science. The search was limited to journal articles and conference proceedings. Librarians were consulted to ensure the comprehensiveness of the literature search. The search applied the PICO framework,[35] using a range of relevant terms to capture all potentially eligible results relating to smartphone mental health interventions for BD (Table S2).

2.2 | Eligibility criteria

Eligible articles should (i) provide data on smartphone-based interventions in patients diagnosed with BD—that could range from symptom monitoring (support-based app) to a wide range of active interventions and could be delivered solely or in part via smartphone devices; (ii) provide at least one outcome on the efficacy or effectiveness of the intervention—including inter-group or pre-post (intra-group) comparisons on number of mood episodes, psychiatric admissions, or clinical scales on (hypo)manic or depressive symptoms, functioning and patient-reported outcome measures (PROM) such as perceived stress, and quality of life. Post hoc analyses of the included studies were also included if they provided relevant information for the analyses that were not available in the original publication.

"Smartphones" were defined as mobile phones with 3G, 4G, or 5G Internet connectivity, along with the ability to download, install and run external applications ("apps"). Eligible studies included randomized controlled trials (RCTs) and observational studies. RCTs using either "inactive" or "active" control groups were eligible for inclusion. "Inactive" control groups were classified as those in which participants received no intervention during the trial period (or were put into a waitlist until pre-and-post measures had been collected from both groups). "Active" control groups were categorized as those which attempted to control for the time and attention given to people in the smartphone intervention condition, by using apps without therapeutic content (placebo-app), in-person interventions, or other forms of activities or patient contact. No restrictions were placed on clinical or demographic characteristics of eligible samples, except for a diagnosis of BD.

Exclusion criteria were (i) interventions not delivered through smartphones (e.g., exclusive phone calls, phone messaging, only SMS, or computer-delivered interventions); (ii) articles not directly concerned to mental health; (iii) reviews, meta-analyses, comments, and letters to the editor. The references of all full-text articles and relevant review articles were also searched for additional studies.

Two independent investigators judged article eligibility (GA and DHM), with any disagreements resolved through discussion by a third investigator (EV).

2.3 | Data extraction

The relevant data from individual articles were extracted using a preconceived and standardized data extraction form. Two

researchers (GA and DHM) were involved in the data extraction process. Information extracted included study characteristics (Table 1), smartphone-app/intervention details (Table 2), outcome measures on the efficacy or effectiveness of the intervention (Tables 3 and 4), and user-engagement indicators (UEIs) (Table 5). Relevant data unavailable in the articles were requested to the authors of the studies.

2.4 | Quality assessment of the included studies

The quality of each manuscript included was independently assessed by two blind researchers (GA and DHM) and disagreements were resolved by consensus in a meeting with another researcher (EV). RCTs were assessed using the Cochrane Collaboration's Risk of Bias tool 2 (RoB 2)[36] (Figure S2). Observational studies were assessed using the National Institutes of Health (NIH) quality assessment tool for before-after (Pre-Post) study without control group[37,38] (Table S3).

2.5 | Statistical analyses

We conducted all analyses in R (R version 4.1.1),[39] using a random-effects model[40] to account for between-study heterogeneity. We used the MetaNSUE package 2.4[41] to unbiasedly include studies that only report that the differences did not reach statistical significance.

The pre-post change scores in depressive and/or (hypo)manic symptom severity, perceived stress, functioning, and quality of life for smartphone intervention and control conditions were pooled to calculate the standardized difference between conditions as Hedge's g with 95% confidence intervals (CI). Heterogeneity between studies was examined using I^2 statistics. A p-value of <0.05 was considered statistically significant, and Bonferroni correction for multiple testing was applied.

3 | RESULTS

The literature search identified 6034 studies (2670 after duplicates were excluded). Of these, 13 articles fulfilled the selection criteria and were included in the qualitative synthesis. Five RCTs were included in the meta-analyses (Figure S1). The included studies consisted of (i) seven RCTs,[42–47] (ii) three observational pre-post studies,[48–50] and (iii) three post hoc pre-post analyses from the included studies.[51–53]

3.1 | Study quality

Five RCTs were at low risk of bias according to the RoB 2 tool (Figure S2). Six RCTs[42–47] were single-blinded (for researchers conducting outcome assessment) and one used a fully remote randomized controlled crossover waitlist trial design. Two RCTs had some risk of bias: one because the scales used to measure affective symptoms were not specific,[46] the other due to an entirely remote design (including self-reported diagnoses, scales, and waitlist bias).[54] All observational studies[48–50] were of fair quality according to the NIH quality assessment tool (Table S3). Most quality concerns were related to blinding, lost to follow-up (and statistical analyses accounting for it), and lack of multiple measures for outcomes.

3.2 | Characteristics of the included studies

The characteristics of the included studies, including design, participants, diagnoses, and study length have been reported in Table 1.

Six RCTs[42–47] were randomized single-blind trials, and one used a fully remote randomized controlled crossover waitlist trial design.[54] Six RCTs included two arms (intervention vs. control groups)[43–45,55] except for one that included three arms (one intervention and two control groups: one active and one inactive).[46] The RCTs included a total of 479 participants with BD (54.5% female; 264 intervention group; 215 controls)—306 with BD type I (BD-I), 64% of the total sample. Only one study did not report the type of the included participants with BD[47] and another included participants with self-reported diagnoses and was not accounted in the reported numbers.[54] Three studies included patients with other diagnoses than BD,[46,47,52,54] but only the data regarding patients with BD (Table 1) were used for the meta-analyses. The pooled mean age was 39.56 years. The duration of RCTs ranged from 10 weeks[45] to 9 months[55] with a median of 25 weeks.[12–26]

The three observational studies included a total of 278 participants with BD (60% female)—54 with BD-I (specified in two studies[48,50]). The duration of the studies ranged between 3[48] and 6 months.[49,50] The pooled mean age of participants was 38.70 years.

3.3 | Type of intervention and smartphone-app characteristics

The characteristics of the intervention and the smartphone apps from each study have been summarized in Table 2.

Smartphone-based interventions ranged from (i) self-monitoring (SM), including a feedback from the clinic[42–44] or from the app only[45]—aimed at identifying prodromal symptoms for intervention—to (ii) app-delivered personalized cognitive-behavioral therapy (CBT) plus SM[46] or (iii) app-delivered interventions including CBT, coping strategies, skills training, psychoeducation, medication reminders plus SM.[47] Apps did not collect passive data except for two studies[42,44] (including phone usage, social activity, physical activity, and mobility). Two apps included direct psychoeducation,[47,54] four of them indirect psychoeducation through feedback loop[42–44] or interventions based on symptoms or early warning signs,[45] and one app did not provide psychoeducative feedback.[46] Four apps included notifications. These were via SMS to engage patients in CBT interventions,[46] to remember survey completion,[45] or via app to prompt patients to perform the daily

TABLE 1 Characteristics of the included studies

Author, year, location	Study length (Weeks)	Diagnoses	Clinical State	Arms	Sample size *N* total[a]	*N* smartphone intervention	Controls	Mean age, Years (SD)[b]	FemaleN (%)[b]	BD type 1 *N* (%)[b]
Randomized control trials										
Faurholt-Jepsen et al.; 2020, Denmark[42]	≥26 (6 months)	BD[c]	Patients included after discharge from hospitalization[d]	2	98	47	51	41.6 (13.3)[e]	51 (52.04)	57 (58.2)
Faurholt-Jepsen et al.; 2019, Denmark[44]	≥39 (9 months)	BD[c]	Any[f]	2	129	85	44	43.0 (12.4)[e]	76 (58.9)	75 (58.1)
Faurholt-Jepsen et al.; 2015, Denmark[43]	≥26 (6 months)	BD[c]	Euthymia to moderate episodes[g]	2	67	33	34	29.1 (7.5)[e]	45 (67.2)	45 (67.2)
Depp et al.; 2019; US[46]	12	BD-I[h] SCZ/SZA	Stable, but with some symptoms[i]	3[j]	57[k]	22+15[l]	20[l]	51.2 (11.5)[m]	113 (49.3)[m]	57 (100)[l]
Depp et al.; 2015; US[45]	24 (10 months of intervention)	BD[h]	Euthymia to moderate episodes[n]	2	82	41	41	47.5 (12.8)	48 (58.5)	36 per arm (87.8)
Ben-Zeev et al.; 2018, US[47]	12	BD[o] MDD SCZ/SZA	Minimal-to-Severe symptoms[p]	2	46 BD[m] (163 total)	21[l]	25[l]	49 (10.1)[m]	67 (41.1)[m]	NR
Ben-Zeev et al.; 2021, US[54]	8.6/12.9 (60/90 days) for active/control groups	BD[r] MDD SCZ/SZA	Minimal-to-Severe symptoms[r]	2	111 BD[m] (315 total)	50[l]	61[l]	37.9 (11.6)[m]	264 (83.8)[m]	NR
Observational studies										
Hidalgo-Mazzei et al.; 2018, Spain[49]	≥26 (6 months)	BD[q]	Any	1	201	201	0	36.6 (11)	127 (63.2)	NR[q]
Hidalgo-Mazzei et al.; 2016, Spain[48]	≥13 (3 months)	BD[c]	Euthymia (HDRS < 8 and YMRS < 6)	1	49	49	0	43.9 (11.4)	21 (42.9)	33 (67.3)
Ryan et al.; 2021, US[50]	≥26 (6 months)	BD[h]	Any[s]	1	28	28	0	44.7 (11.2)	19 (67.9)	21 (75)

Abbreviations: BD, Bipolar Disorder; FAST, Functioning Assessment Short test; HDRS-17, Hamilton Depression Rating Scale with 17 items version; MADRS, Montgomery and Asberg Depression Scale; NS, Not Specified; SM, Self-Monitoring; TAU, Treatment as Usual; YMRS, Young Mania Rating Scale.

[a]Total number of patients analyzed, not allocated.

[b]Characteristics of all participants included, except otherwise specified.

[c]ICD-10 criteria.

[d]Hospital discharge following an affective episode (depression, mania, or mixed episode). Depression and mania symptoms scales at baseline ranged from minimal to moderate. However, the Functional Assessment Short Test (FAST) at baseline was superior to 30 in both groups, thus indicating functional impairment.

[e]Data for the intervention group only.

[f]*No exclusion criteria regarding clinical state of patients. Included participants rated* <=3 on one of three items constituting the domination by symptoms factor from the Recovery Assessment Scale (RAS), indicating a need for the type of resources both interventions may offer.

[g]Hamilton Depression Rating Scale (HAMD-17) and Young Mania Rating Scale (YMRS) scores ≤ 17.

[h]DSM-IV criteria.

[i]Outpatients prescribed stable psychotropic medication regimens for the prior 3 months. Patients rated greater than 3 (1 = not present to 7 = extremely) on at least one of the Brief Psychiatric Rating Scale (BPRS) items 3 (Depression), 7 (Elevated Mood), 10 (Hallucinations), or 17 (Emotional Withdrawal).

TABLE 1 (Continued)

[j](1) CBT2go; (2) Self-Monitoring (SM); (3) Treatment as Usual (TAU). (1) CBT2go and (2) SM were active arms, (3) TAU was a control arm.

[k]*Only participants with BD-I. CBT2go included 22 participants, SM 15 and TAU 20; a total of 57 participants with BD-I. The total sample consisted of* 255 participants, 85 per arm *including patients with BD, SCZ, and SZA.*

[l]Data specific to participants diagnosed with BD.

[m]*Data not specific to the population with BD, but to the whole sample including patients with other diagnoses (BD, SCZ, and SZA in Depp 2019*[46]*; BD, MDD, SCZ, and SZA in Ben-Zeev 2018,*[47] *and Ben-Zeev 2021*[54]*).*

[n]From euthymia to moderate episodes (excluded patients with MADRS > 32 and YMR > 20).

[o]Diagnostic criteria were not specified and diagnoses were recorded from the electronic health records.

[p]*No exclusion criteria regarding clinical state of patients. Included participants rated* <=3 on one of three items constituting the domination by symptoms factor from the Recovery Assessment Scale (RAS), indicating a need for the type of resources both interventions may offer.

[q]Previously diagnosed with BD by a mental health professional. Upon inclusion, participants were asked to complete an online modified version of the Mood disorder questionnaire (MDQ) and were included if they met eligibility criteria.

[r]Self-reported diagnoses and symptoms. No exclusion criteria regarding clinical state of patients.

[s]Self-reported symptoms at baseline, 3, and 6 months.

evaluations[42] or daily training.[54] In three studies, participants were contacted by text message, phone call, or e-mail if there were signs of deterioration of depressive or manic symptoms.[42–44] The number of sessions per day ranged from one[42–44] to three[46,47] with up to 11 items evaluated per session.[45] Six RCTs offered the option to receive loaned phones, with only three allowing the use of personal phones also[42,44,46] and four RCTs included economic compensations for participants[45–47] ranging from 25$[45] to 50$[46] per session/assessment, or the possibility to win up to 1000$ in lottery after completing assessments.[54]

Two of the included observational studies used a smartphone-based psychoeducational program for bipolar disorder (SIMPLe) that included SM through ecological momentary assessments (EMA) and provided adapted psychoeducation through messages according to the clinical states, potential relapses, and risk situations. The first study was conducted with the original version (1.0),[48] in a clinical setting, and the second study with the upgraded version (1.5)[49] that incorporated some features including medication reminders, personalized prodromal symptoms register, and gamification modules to enhance engagement. This second study was conducted in a web-based 100% online setting worldwide. The SIMPLe app did not collect passive data, included app notifications through reminders of pending tests (version 1.0) and additional medication reminders (version 1.5). Only one session per day was required with five items assessed. Only personal phones could be used and no economic compensation was offered. Another observational study[50] used an app that included brief modules to provide self-management activities for managing everyday needs of individuals with BD. The app did not collect passive data and users received economic compensation for use.

3.4 | Efficacy and effectiveness of smartphone-based interventions

The outcomes on efficacy and effectiveness of smartphone-based interventions have been summarized in Table 3.

We performed seven meta-analyses including five RCTs comparing the pre-post change scores in depressive and (hypo)manic symptoms, functioning, quality of life, and perceived stress between smartphone interventions and control conditions (Table 4). There was considerable heterogeneity among studies and no meta-analysis reached statistical significance. Results were also inconclusive regarding affective relapses and psychiatric readmissions. However, a non-significant tendency toward better outcomes in the intervention group was found in the meta-analyses for all variables except in the meta-analyses including participants with higher baseline affective symptoms (Figure 1).

Four smartphone-based interventions from the included RCT did not show efficacy in reducing depression outcomes,[42–44,47] and three did: one study included only participants with BD in which the comparator was paper-and-pencil monitoring,[45] another study measuring depressive outcomes with the brief psychiatric rating scale and only

TABLE 2 Intervention/App characteristics

Author, year, location	App – Study reference	Version	Type of intervention	Phone	Notifications (Type)
Randomized control trials					
Faurholt -Jepsen et al.; 2020, Denmark[42]	Monesnso system	Upgraded	(1) CBT+SM[a] (2) Control smartphone[a]	Personal or loaned	Yes[b]
Faurholt -Jepsen et al.; 2019, Denmark[44]	Monesnso system (MONARCA II)	Upgraded	(1) SM[a] (2) Control smartphone[a]	Personal or loaned	No[c]
Faurholt-Jepsen et al.; 2015, Denmark[43]	Monesnso system (MONARCA I)	Original	(1) SM[a] (2) Control smartphone[a]	Loaned	No[c]
Depp et al.; 2019; US[46]	MOBIT[e]	Original	(1) CBT + SM[f] (2) SM only[f] (3) TAU	Personal or loaned	Yes (SMS[g])
Depp et al.; 2015; US[45]	PRISM[j]	Original	(1) SM[j] (2) Paper-and-pencil monitoring	Loaned	Yes (SMS[k])
Ben-Zeev et al.; 2018, US[47]	FOCUS	Original	(1) CBT, coping strategies, skills training, psychoeducation + Self-management + SM[m] (2) Self-management + SM[m]	Loaned	No
Ben-Zeev et al.; 2021, US[54]	CORE	Original	(1) Cognitive training to reduce dysfunctional self-talk and increase resilience[q] (2) Waitlist (1-month) to receive app.	Personal	Yes[r]
Observational studies					
Hidalgo-Mazzei et al.; 2018, Spain[49]	SIMPLe[n]	1.5 (Updated)	Smartphone-based psychoeducational program for BD[o]	Personal	Yes (reminding of pending test + medication reminder)
Hidalgo-Mazzei et al.; 2016, Spain[48]	SIMPLe[p]	1.0 Original	Smartphone-based psychoeducational program for BD[n]	Personal	Yes (reminding of pending test)
Ryan et al.; 2021, US[50]	Life Goals app[s]	Original	Self-management skills and goal-setting program for BD[t]	Personal	Yes (reminders of module completion, not logging in, and self-evaluation tests)

Abbreviations: CBT, Cognitive behavioral therapy; SM, Self-Monitoring; TAU, Treatment as Usual.

[a]*(1) MONARCA: daily electronic self-monitoring using smartphones including a clinical feedback loop between the patients and the clinic, where the self-monitored data were sent to the clinic allowing for the study nurse to review the data and contact the patients if there were signs of deterioration, thereby allowing for intervention on prodromal depressive and manic symptoms, and a feedback loop within the patients themselves, where the self-monitored data were visualized graphically to the patients providing them with an overview of the entered data, thereby providing possibilities for an increased illness insight and understanding. The latest Monsenso System version (2020 RADMIS trial[42]) included CBT modules with cognitive restructuring, rumination-focused CBT, and psychoeducation. (2) Control group: use of a smartphone for normal communicative purposes.*

[b]Patients were prompted daily by an alarm in the Monsenso system at a self-chosen time during the day by the MONARCA system to evaluate the following subjective items: mood, sleep length, medication taken, activity, irritability, mixed mood, cognitive problems, alcohol consumption, stress, menstruation for women, and individualized early warning signs.

[c]However, if there were signs of deterioration of depressive or manic symptoms, patients were contacted by text message, phone call, or e-mail.

[d]Economic costs due to data traffic were refunded.

[e]A web-based program called Mobile Online Behavioral Intervention Technology (MOBIT) delivered interactive surveys to the device that contained elements personalized from the individual session. In addition to SM, the active group received single-session in-person intervention called CBT2go augmented by mobile interactions.

Psychoeducation (Type)	Collection of passive data (Type)	Feedback	Sessions in a day	Number of items evaluated in a session	Economic compensation
Yes (CBT modules including psychoeducation, cognitive restructuring, and rumination-focused CBT)	Yes (Phone usage, Social activity, Physical activity, Mobility)	Clinical and graphical feedback loop	1	10[b]	No
Not directly (but indirect feedback loop)	Yes (Phone usage, Social activity, Physical activity, Mobility)	Clinical and graphical feedback loop	1	10[b]	No[d]
Not directly (but indirect feedback loop)	No	Clinical and graphical feedback loop	1	10[b]	No[d]
No (only in in-person sessions)	No	Messages of personalized CBT intervention on maladaptative beliefs + Graphical feedback review	3[h]	2–5[i]	Yes (50$ per session)
Not directly (indirect interventions[f])	No	Graphical feedback on the mood course	2	11[l]	Yes ($25 for each completed assessment)
Yes (through the interventions, specifically for the "sleep" and "medication" interventions[m]	No	Graphical feedback on symptoms course+Weekly calls from mHealth support specialist providing technical and clinical assistance	3	NR	Yes (30$ per assessment)
Yes (about maladaptative beliefs)	No	Users can endorse or discard statements, add their personal toolbox and access it anytime. + Graphical feedback is provided.	Daily training not quantified (total of 53 modules)	NR	Yes (possibility to win up to 1000$ in lottery after completing assessments)
Yes (direct psychoeducative messages)	No	Graphical feedback + Psychoeducative messages	1[o]	5[p]	No
Yes (direct psychoeducative messages)	No	Graphical feedback + Psychoeducative messages	1[o]	5[p]	No
Yes (through specific modules[t])	No	Feedback on symptoms experience. Not specified feedback on mood/symptoms or according to modules.	Ideally one module per day	NR	Yes (for every day using the app and for each survey completed: up to $210 for the whole study)

TABLE 2 (Continued)

[f](1) CBT2go; (2) Self-Monitoring (SM). The CBT2go and SM arms included a 90-minute session that included psychoeducation about the diagnosis, causes, symptoms, and treatments for mental illness, and the importance of self-monitoring symptoms. After the 90-minute session, the SM smartphone-based intervention included assessment contents only (about the frequency/severity of symptoms, socialization, and medication adherence), without any of the intervention content. The CBT2go smartphone-based intervention included the aforementioned assessment contents plus intervention content that branched from each of the maladaptive beliefs to offer a potential alternative or adaptive belief, personalized by the individual in the in-person session, accompanied by adaptive behaviors. In the TAU arm, participants only completed face-to-face assessments at baseline and 6-, 12-, and 24-week follow-up.

[g]The mobile device prompted participants to engage in cognitive restructuring.

[h]Mood or Voices algorithms were delivered during the morning survey, Socialization in the afternoon, and Medication Adherence in the evening.

[i]The first question pertained to symptom severity/frequency, socialization, and medication adherence and the second to the presence of one of 3–4 current maladaptive beliefs corresponding to that domain. The intervention content branched from each of the maladaptive beliefs to offer a potential alternative or adaptive belief, personalized by the individual in the in-person session, accompanied by adaptive behaviors.

[j](1) Personalized Real-Time Intervention for Stabilizing Mood (PRISM): Augmentative mobile ecological momentary intervention targeting self-management of mood symptoms. The app provided pre-programmed interactive algorithm-based responses based on symptoms or early warning signs reported. Participants were presented with a random selection of one of the implementation intentions that they had previously developed. They were also asked if they wanted to see a different coping strategy drawn at random from a list. (2) Paper-and-pencil monitoring. Both groups received four sessions of individual therapy to identify an "Action Plan" which specifies adaptive responses to depressive and manic symptoms as well as early warning signs and triggers of illness exacerbations.

[k]Only to remember survey completion.

[l]At each survey, participants answered four questions: What are you doing?, Where are you?, and Who are you with? Subsequently, respondents answered six questions about mood state. Next, participants rated their current mood state on a 9-point bipolar anchored scale. Depending upon their rating on the 9-point mood scale, participants were presented with a random selection of one of the implementation intentions that they had previously developed. They were also asked if they wanted to see a different coping strategy drawn at random from a list. On the Evening survey, participants were first asked if they had tried the strategy that was recommended earlier in the day and whether it was helpful.

[m](1) Active group: Smartphone-delivered intervention (FOCUS). The system includes preprogrammed daily self-assessment prompts and on-demand functions that can be accessed 24 hours a day. Self-management content targets five broad domains: voices (coping with auditory hallucinations via cognitive restructuring, distraction, and guided hypothesis testing), mood (managing depression and anxiety via behavioral activation, relaxation techniques, and supportive content), sleep (sleep hygiene, relaxation, and health and wellness psychoeducation), social functioning (cognitive restructuring of persecutory ideation, anger management, activity scheduling, and skills training), and medication (behavioral tailoring, reminders, and psychoeducation). Content can be accessed as either brief video or audio clips or sequences of digital screens with written material coupled with images. (2) Control group: Clinic-based group intervention (Wellness Recovery Action Plan [WRAP]). WRAP is a group self-management intervention led by trained facilitators with lived experience of mental illness. Sessions follow a sequenced curriculum, and specific group discussion topics and examples draw from the personal experiences of the participants and cofacilitators. The model emphasizes individuals' equipping themselves with "personal wellness tools," each focusing on recovery concepts (e.g., hope, personal responsibility, and self-advocacy), language (e.g., person-first recovery language), development of a WRAP (e.g., establishing a daily maintenance plan and identifying and responding to triggers and early warning signs), and encouraging positive thinking (e.g., changing negative thoughts to positive thoughts, building self-esteem, suicide prevention, and journaling). Facilitators incorporate these tools into a written plan, which includes daily maintenance, identification of triggers and methods to avoid them, identification of warning signs and response options, and a crisis management plan.

[n]Smartphone-based psychoeducational program for bipolar disorder (SIMPLe 1.5, improved version of SIMPLe 1.0). In addition to SIMPLE 1.0 functions, the following new features and characteristics were incorporated in SIMPLe 1.5: Medication reminders, Personalized prodromal symptoms register and self-monitoring, Gamification modules to enhance engagement, Psychoeducational messages community, Personalized configuration of local emergency services telephone number and e-mail reference of the patient's preference.

[o]Plus weekly complementary session.

[p]Smartphone-based psychoeducational program for bipolar disorder (SIMPLe 1.0) aimed to collect ecological momentary assessments (EMA) to adapt psychoeducational messages according to the clinical states, potential relapses, and risk situations. Included features in SIMPLe 1.0: Daily short graphic 5-item screening test (i.e., mood, energy, sleep time, medication adherence, and irritability); Weekly, YES or NO test, considering all DSM-5 criteria for manic and depressive episodes including suicide thoughts; System notifications pop-up in the smartphone reminding of a pending test at the time configured by the user; If there was an emergency detected by the application (i.e., suicide ideation), a treating psychiatrist was alerted, in order to contact the patient.

[q]The CORE intervention is a system designed to increase the cognitive flexibility of individuals struggling with a range of mental health problems through brief daily training. CORE was specifically designed to help counteract dysfunctional thoughts in multiple domains that are relevant to the subjective experience of having several mental illness. The app included modules focused on countering maladaptive beliefs in common mental health domains (e.g., self-esteem, distinguishing thoughts from feelings, social anxiety, and catastrophizing) as well as modules (e.g., stigma related to mental illness, threat perception and persecutory ideation, hopelessness, strength-based recovery, self-care, and treatment seeking).

[r]Push notifications reminded users to complete their daily training.

[s]The Life Goals app is based on the Life Goals Collaborative Care (LGCC) program. This program provides proactive care for patients through patient self-management education, care coordination across providers, and decision support tools for providers. LGCC focuses on empowering patient-self management skills with a series of 6 or more self-management sessions customized to individual needs and focused on mental and physical wellness, understanding symptoms, and setting personal goals.

[t]The Life Goals app includes succinct, 5- to 10-minute-long modules (a total of 13) that provide self-management activities for managing the everyday needs of individuals with BD, including mood symptom coping strategies, stigma concerns, emotional self-awareness and family support, anger and irritability, and preparation for doctor's visits. The content and wording were adapted from the original LGCC program.

when comparing the intervention with treatment-as-usual (TAU), but not when comparing the intervention with the active control group,[46] the third RCT that showed efficacy used a fully remote randomized controlled crossover waitlist trial design with self-reported diagnoses and results on depression were including other diagnoses.[54]

None of four[42–45] of the smartphone-based interventions from the included RCTs showed efficacy in reducing manic symptoms. Only one[54] of seven RCTs showed efficacy in improving function outcomes. Quality of life was improved in one[44] of four RCTs, and perceived stress was reduced in two[42,44] of three RCTs (Table 3; Figure 1).

Regarding affective relapses, one RCT[44] showed a reduction in manic relapses, but an increased risk of depressive relapses, whereas another study did not show significant differences in affective relapses of any polarity.[42]

Regarding psychiatric readmissions, one RCT[42] did not show reduction in rate or duration of readmissions.

3.5 | Baseline affective symptoms

Participants' baseline affective symptoms ranged from absent-to-minimal in three studies,[42–44] mild symptoms in one study,[45] mild-to-severe depressive symptoms in two studies[47,54] including participants with major depressive disorder, schizophrenia and schizoaffective disorder, and mild-to-moderate general psychiatric symptoms in another study (including participants with schizophrenia and schizoaffective disorder).[46]

Considering the severity of baseline depression symptoms with the Hamilton Depression Rating Scale (HDRS) ≥7 at baseline, participants in the intervention group experienced higher levels of depressive symptoms compared with the control group in two[42,43] of three RCTs. Considering the severity of baseline manic symptoms with the Young Mania Rating Scale (YMRS) ≥7 at baseline, participants in the intervention group experienced higher levels of manic symptoms compared with the control group in two[42,44] of three RCTs. When assessing pre-post improvements considering the severity of baseline depression, participants with moderate-to-severe depression at baseline showed significantly marked reductions compared to participants with mild depression.[47]

Regarding pre-post improvements between baseline and the end of the smartphone-based interventions, three studies revealed favorable effectiveness outcomes.[46,47,54]

3.6 | Active or inactive control groups

Active control groups varied widely, ranging from the use of a smartphone for communicative purposes,[42–44] clinic-based group self-management interventions,[47] smartphone self-monitoring plus face-to-face psychoeducation sessions,[46] to paper-and-pencil monitoring.[45] One study used a waitlist control group and a fully remote randomized crossover design,[54] which clearly differs from the design of the other RCTs.

On the one hand, when assessing the efficacy of smartphone-based active psychological-validated interventions (i.e., CBT or psychoeducation)—not SM only—one of the included studies assessing a smartphone-based CBT intervention compared it to smartphone-based SM-only and to usual clinical assessments without smartphone.[46] However, another study, compared an app including CBT, coping strategies, psychoeducation, and SM to a clinically validated self-management group intervention (Wellness Recovery Action Plan [WRAP]).[47] In this latter study, both the smartphone psychological intervention and the clinically validated intervention were recovery-oriented, used an array of empowerment and self-management techniques, and involved similar intervention periods; with evidence suggesting that both interventions were engaging and beneficial to people with serious mental illness.[56,57] The differences between these approaches represent the core distinctions between mHealth and clinic-based models of care (i.e., accessed in one's own environment vs. administered in a center, largely automated vs. person delivered, and on demand vs. scheduled).

On the other hand, when assessing the efficacy of a smartphone-based interventions consisting of SM (including or not app/clinician feedback),[43–46] one study compared it to paper-and-pencil monitoring,[45] two studies to the use of a smartphone only for communicative purposes (without SM),[43,44] and another study to usual clinical assessments without smartphone.[46]

3.7 | Observational studies

Two observational studies showed effectiveness of the smartphone-based interventions: one in reducing depressive and manic symptoms, as well as in improving medication adherence and biological rhythms,[51] between baseline and the end of the intervention, especially on participants with more use of the smartphone-based intervention,[48] and the other in improving wellbeing, self-perceptions of disease and functioning.[49] One study included only euthymic BD patients[48] and the other participants at any phase of the disease, but baseline depressive or manic symptoms were not registered.[49] No sub-analyses were made in people with higher baseline depressive or manic symptoms.

Another observational study[50] did not show significant changes in self-reported affective symptoms or functioning when comparing baseline assessments to 3- or 6-month follow-up assessments. The study had a small sample size and it did not find significant relationships between app use measures and changes in outcome measurements.

3.8 | User engagement indicators

UEIs of the smartphone-based intervention from each study have been summarized in Table 5.

TABLE 3 Efficacy and effectiveness of the smartphone-based intervention in BD

Author, year, location	Type of intervention	Comparison	Study length (Weeks)	Depression score at baseline (SD) or [IQR]	Manic score at baseline (SD) or [IQR]	Follow-up Assessments	Type of assessment	Depression outcome
Efficacy of the intervention in Randomized Control Trials								
Faurholt-Jepsen et al.; 2020, Denmark[42]	(1) SM[b] (2) Control smartphone[b]	1 vs. 2	≥26 (6 months)	HDRS-17: 6 [2–12][a]	YMRS: 3 [0–7][a]	Baseline, 3 and 6 months	Self-reported + clinician-administered	HDRS-17
Faurholt -Jepsen et al.; 2019, Denmark[44]	(1) SM[b] (2) Control smartphone[i]	1 vs. 2	≥39 (9 months)	HDRS-17: 6 [4-11][a]	YMRS: 2 [0-5][a]	Baseline, 4 weeks, 3,6,and 9 months	Self-reported + clinician-administered	HDRS-17

Effect size	*p*-value	Mania outcome	Effect size	*p*-value	Function outcome	Effect size	*p*-value	Other outcomes: quality of life, perceived stress, mood episodes prevention, the moderator effect of baseline symptoms, and pre-post differences in outcome
−0.11 [−2.5; 2.29][c]	0.93	YMRS	1.89 [0.008; 3.78][c]	**0.05**	FAST	−3.03 [−7.43; 1.37][c]	0.18	**Quality of life**: *WHOQoL-BREF*: 0.61 [−3.99; 5.2][c]; $p = 0.80$ **Perceived stress**: *PSS*: −7.18 [−13.5; −0.86][c]; $\mathbf{p = 0.026}$ **Differences in rate of relapses**: Depressive episodes: 0.55 [0.16; 1.89][d]; $p = 0.34$ Manic episodes: 0.87 [0.12; 6.25][d]; $p = 0.89$ Affective episodes: 0.63 [0.29; 1.37][d]; $p = 0.24$ **Psychiatric readmission**: 1.05 [0.52; 1.85]; $p = 0.88$ **Severity of baseline depression symptoms** [c]: Participants in the intervention group with HDRS ⩾7 at baseline experienced higher levels of depressive symptoms compared with the control group (adjusted model: $B = 2.55$, 95% CI 0.08– 5.02, $\mathbf{p = 0.05}$) ($n = 34$). **Severity of baseline manic symptoms** [c]: Participants in the intervention group with YMRS ⩾7 at baseline experienced higher levels of manic symptoms compared with the control group (adjusted model: B = 4.17, 95% CI 1.07–7.27, $\mathbf{p = 0.008}$) ($n = 20$).
0.61 [−0.77; 2.0][c]	0.38	YMRS	−0.25 [−1.1; 0.59][c]	0.56	FAST	0.36 [−2.5; 3.2][c]	0.81	**Quality of life**: *WHOQOL-BREF*: 4.00 [0.04; 7.97][c]; $\mathbf{p = 0.048}$ **Perceived stress**: *PSS*: −2.08 [−4.15; −0.01][c]; $\mathbf{p = 0.049}$ **Differences in rate of relapses**: smartphone-based monitoring may reduce the risk of relapse of manic episodes but increase the risk of relapse of depressive episodes. Depressive episodes: 2.89 [1.02; 8.23][d]; $\mathbf{p = 0.047}$ Manic episodes: 0.17 [0.037; 0.78][d]; $\mathbf{p = 0.022}$ Affective episodes: 0.76 [0.41; 1.42][d]; $p = 0.39$ **Severity of baseline depression symptoms** [c]: When including patients with HDRS ⩾7 at baseline, there was no difference between the (1) and (2) (adjusted model: B = 0.74, 95% CI −1.30 to 2.79, p = 0.48) ($n = 64$). **Severity of baseline manic symptoms** [c]: Participants in the intervention group with YMRS ⩾7 at baseline experienced higher levels of manic symptoms compared with the control group (adjusted model: B = 4.21, 95% CI 0.45–7.97, $\mathbf{p = 0.028}$) ($n = 19$). **Excluding mixed depressive symptoms** [c] Excluding mixed depressive symptoms, patients in the intervention group experienced statistically significantly lower levels of manic symptoms compared with the control group (adjusted model: B = −1.11, 95% CI −2.22 to −0.01, $\mathbf{p = 0.050}$) ($n = 88$).

TABLE 3 (Continued)

Author, year, location	Type of intervention	Comparison	Study length (Weeks)	Depression score at baseline (SD) or [IQR]	Manic score at baseline (SD) or [IQR]	Follow-up Assessments	Type of assessment	Depression outcome
Faurholt-Jepsen et al.; 2015, Denmark[43]	(1) SM[b] (2) Control smartphone[b]	1 vs. 2	≥26 (6 months)	HDRS-17: 9 [4–16][a]	YMRS: 2 [0–7][a]	Monthly	Self-reported + clinician-administered	HDRS-17
Depp et al.; 2019; US[46]	(1) Active intervention + SM (2) SM only (3) TAU	1 vs. 2 (active vs. active)	12	BPRS-14[f,g,h]: 39.3 (6.8)	NR[g]	baseline 6-, 12-, 24-week	Self-reported + clinician-administered	BPRS-14[g]
		2 vs. 3						BPRS-14[g]
		1 vs. 3 (active vs. TAU)						BPRS-14[g]
Depp et al.; 2015; US[45]	(1) SM[j] (2) Paper-and-pencil monitoring	1 vs. 2	10	MADRS: 11.7 (9)[f]	YMRS: 7.4 (6)[f]	Baseline, 6-, 12-, 24-week	Self-reported + clinician-administered	MADRS

Effect size	*p*-value	Mania outcome	Effect size	*p*-value	Function outcome	Effect size	*p*-value	Other outcomes: quality of life, perceived stress, mood episodes prevention, the moderator effect of baseline symptoms, and pre-post differences in outcome
2.02 [−0.13; 4.17][c]	0.066	YMRS	−0.34 [−1.14; 0.47][c]	0.41	FAST	0.96 [−4.36; 6.28][c]	0.72	**Quality of life**: *WHOQOL-BREF*: −1.24 [−5.18; 2.7][c]; $p = 0.54$ **Perceived stress**: *PSS*: 2.40 [−0.33; 5.13][c]; ***p* = 0.085** **Severity of baseline depression symptoms**: When including patients with HDRS >7 at baseline, patients in the intervention group experienced significantly more depressive symptoms compared to the control group (adjusted model: B = 2.69, 95% CI 0.001–5.37, ***p* = 0.049**) ($n = 38$). **Severity of baseline manic symptoms**: Participants in the intervention group with YMRS >7 at baseline experienced significantly fewer manic symptoms compared to the control group (adjusted model: B = −6.32, 95% CI −9.21 to −3.34, **p<0.001**) ($n = 13$). **Excluding mixed depressive symptoms** Excluding mixed depressive symptoms, patients in the intervention group experienced borderline significantly fewer manic symptoms compared to the control group in the adjusted model (B = −1.07, 95% CI −2.15 to 0.005, $p = 0.051$) ($n = 45$). **Post hoc analysis**[53]: Patients with BD-I compared to BD-II experienced higher mean level of mood on a scale from −3; +3, more time euthymic and less time with depressive symptoms. Level of mood −0.19 vs. −0.54[e]; $p = 0.02$ % time euthymic 74.5% vs. 51%[e]; $p = 0.03$ % time with depressive symptoms 18.8% vs. 45.1%[e]; $p = .01$
<0,01[h]	0.685	-	-	-	SLOF	0.12[h]	0.353	**Pre-post differences in primary outcome (BPRS-14; cohen's d; p)**[h] (1) 0,36; <0,001 (2) 0,26; 0,005
Cohen's d: 0.2[i]	NR	–	–	–	SLOF	0.24[h]	0.383	**Treatment response (25 % of improvement in BPRS-14 score; NNT)**[h] 15,6 vs. 9,6% (NNT = 15.6)
Cohen's d: **1.24**[i]	**NR**	–	–	–	SLOF	0.3[h]	0.063	**Treatment response (25 % of improvement in BPRS-14 score; NNT)**[h] 21,1 vs. 9,6% (NNT = 8.7)
Cohen's d: 0.48	**0.042**	YMRS	Cohen's d: 0.33[k]	NS	IIS[l]	Cohen's d: 0.18	NS	–

TABLE 3 (Continued)

Author, year, location	Type of intervention	Comparison	Study length (Weeks)	Depression score at baseline (SD) or [IQR]	Manic score at baseline (SD) or [IQR]	Follow-up Assessments	Type of assessment	Depression outcome
Ben-Zeev et al.; 2018, USA[47]	(1) CBT, coping strategies, skills training, psychoeducation + SM[m] (2) Self-management intervention[m]	1 vs. 2	12	BDI-II[a,h]: (1): 22 (11.2); (2) 19.5 (12.1)	NR	Baseline, 12 weeks,and 6 months	Self-reported + clinician-administered	BDI-II
Ben-Zeev et al.; 2021, US[54]	(1) Cognitive training to reduce dysfunctional self-talk and increase resilience[q] (2) Waitlist (1-month) to receive app.	1 vs. 2	8.6/12.9 (60/90 days) for active/ control groups	BDI-II[a,h]: (1): 33.5 (13.7); (2) 34.6 (13.3)	NR	Baseline (T1), 30 days (T2), 60 days (T3) after the app was given[q]	Self-reported	BDI-II

Author, year, location	Type of intervention	Comparison	Study length (Weeks)	Depression score at baseline (SD) or [IQR]	Manic score at baseline (SD) or [IQR]	Follow-up Assessments	Type of assessment	Depression outcome
Effectiveness of the intervention in observational studies								
Hidalgo-Mazzei et al.; 2016, Spain[48,51]	Smartphone-based psychoeducation (SIMPLe)	-	≥13 (3 months)	HDRS-17: 3.18 (2.69)	YMRS: 2.14 (2.63)	Baseline and 3 months	Self-reported + clinician-administered	HDRS-17
Hidalgo-Mazzei et al.; 2018, Spain[49]	Smartphone-based psychoeducation (SIMPLe)	–	≥26 (6 months)	NR	NR	Before using the app (baseline) and after 6 months	Self-reported	–
Ryan et al.; 2021, US[50]	Life Goals app[r]	–	≥26 (6 months)	PHQ-9[s]: 9.19 (6.99)	ISS[s]: 116.7 (144.7)	Baseline, 3 and 6 months	Self-reported	PHQ-9

Effect size	*p*-value	Mania outcome	Effect size	*p*-value	Function outcome	Effect size	*p*-value	Other outcomes: quality of life, perceived stress, mood episodes prevention, the moderator effect of baseline symptoms, and pre-post differences in outcome
NR[h,n]	NS[h,n]	–	–	–	RAS[n,o]	NR[h,n]	NS[h,n]	**Quality of life**[h,n]: No significant within-group differences between baseline and 12 weeks. **Pre-post improvement between baseline and 3 months for**: **BDI-II**[h]: (1) FOCUS (**−2.76 ±1.09, $p = 0.01$**) and (2) WRAP (−2.33±1.10, $p = 0.03$) **RAS**[h,o]: (1) FOCUS (NS) (2) WRAP (2.44±1.10, $p = 0.03$). **Severity of depression (post hoc analysis[52])**[h]: Participants with **moderate-to-severe depression (20–63 BDI-II)** did have significant reductions in depression symptoms at post-treatment (difference=−4.58; $t = -3.20$; $p = 0.003$) that were also maintained at follow-up (difference=−6.57; t42=−4.20; $p < 0.001$) in contrast to Participants with minimal-mild depression (0–19 BDI-II), who did not have significant reductions in depression symptoms from baseline to post-treatment (difference=−0.22; t31=−0.20; $p = 0.84$) or follow-up.
Cohen's d: 0.58[h]	<0.01[h]	–	–	–	RAS/SDS[h,o]	Cohen's d: 0.61/0.44[h]	0.01[h]	**Pre-post improvement using a series of 1×3 repeated-measures ANOVA between T1, T2 and T3**[h,q]: Significant within-group differences for: **BDI-II**[h]: $F_{2,306}$=43.59; $p < 0.001$ **RAS/SDS**[h,o]: RAS ($F_{2,306}$=65.22; $p < 0.001$) SDS ($F_{2,306}$=56.21; $p < 0.001$). Effect sizes between T1 and T2: BDI-II (d=0.60), RAS (d=0.70), and SDS (d=0.63). All improvements were maintained between T2 and T3.

Effect size	*p*-value	Mania outcome	Effect size	*p*-value	Function outcome	Effect size	*p*-value	Other outcomes
3,18 pre vs. 3,76 post ($t = -0.712$)	0.48	YMRS	2,14 pre vs. 1,51 post ($t = -0.954$)	0.35	NA	NA	NA	**Post hoc analysis**[51]: **BRIAN**[p]: 35,89 pre vs. 31,18 post ($t = 4.29$; $p = <0.001$) **BRIAN *between completers and non-completers (<3 months of app-use)***: 29,47 vs. 35,92 post ($t = 4.29$; $p = <0.001$) ***Medication adherence (4-item Morisky Green score)***: *35,9 pre vs. 31,18 post ($Z = -3.31$; $p < 0.0001$)*
–	–	–	–	–	–	–	–	***WHO-5***: *-3.88; $p < 0.001$* ***SF-36***: *significant improvements in general health perceptions, mental health, physical, emotional, and social role functioning, and bodily pain.* ***Number of hospitalizations***: *NS* ***Number of suicide attempts***: *NS* *No statistical differences in clinical outcomes between completers and non-completers (<5 months of app-use)*
9.19 pre vs. 9.12 post ($t = -.2$)[s]	0.84[s]	ISS	116.7 pre vs. 125.2 post ($t = -0.042$)[s]	0.97[s]	WHO-DAS 2.0[s]	26.6pre vs. 19.0 post ($t = -0.85$)[s]	0.41[s]	**Quality of life**[s]: SF-12 MCS: 39.8 pre vs. 38.7 post ($t = 0.812$; $p = 0.42$)[s]

TABLE 3 (Continued)

Abbreviations: BDI-II: BDI, Beck's Depressive Inventory second edition; BPRS-24, Brief Psychiatric Rating Scale–expanded version; FAST, Functioning Assessment Short test; HDRS, Hamilton Depression Scale; HR, Hazard Rate; MDD, major depressive disorder; ISS, The Internal State Scale; NR, Non-reported; NS, non-significant; PHQ-9, Patient Health Questionnaire – 9; PSS, Cohen's Perceived Stress Scale; RAS, Recovery Assessment Scale; SCZ, schizophrenia; SDS, Sheehan Disability Scale; SF-12 MCS, 12-item Short Form Health Survey Mental Health Component Score; SF-36, short-form health survey; SLOF, Specific Level of Functioning Scale; SM, Self-Monitoring; TAU, Treatment as Usual; WHO-5, the world health organization-five well-being index; WHO-DAS 2.0, World Health Organization Disability Assessment Schedule 2.0; WHOQOL-BREF, The World Health Organization Quality of Life – short version; YMRS, Young Mania Rating Scale.

[a]Baseline data for the intervention group.

[b](1) MONARCA: daily electronic self-monitoring using smartphones including a clinical feedback loop. (2) Control group: use of a smartphone for normal communicative purposes.

[c]Estimated differences in outcome [95% confidence interval]. Adjusted for outcome variable at baseline, age, gender, and HDRS-17 and YMRS scores unless otherwise specified. The MONARCA I study also adjusted for previous hospitalization (yes/no) unless otherwise specified.

[d]Differences in rates of relapse of episodes (Defined as: Depressive HDRS-17 ≥ 13; Manic YMRS ≥ 13) during the study investigated using survival analysis Hazard Rate (HR) [95% confidence interval].

[e]Comparison between 13 BD-I and 20 BD-II (*n* total = 33) in 310 days of follow-up with MONARCA I [IQR 189; 437]. Data presented summarize over 8500 observations. (Outcome: comparison BD-I vs. BD-II; p).

[f]Baseline data for the whole population in the study.

[g]This study did not report specific outcomes on manic and/or depressive symptoms. The BPRS-14 scale used evaluated mostly depressive (9 of 18 items) and psychotic symptoms (4 items). Therefore, we reported the outcomes in the "depression outcomes" section. Only one item evaluated manic symptoms (elevated mood).

[h]*Results are not specific to the population with BD, but to the whole sample including patients with other diagnoses (BD, SCZ, and SZA in Depp 2019*[46]*; and BD, MDD, SCZ, and SZA in Ben-Zeev 2018,*[47] *and Ben-Zeev 2021*[54]*). These results were therefore not included in the meta-analyses. In the study by Ben-Zeev 2021,*[54] inter-group comparisons are between T1 and T2.

[i]Results specific to the population with BD, not to people with other diagnoses included in the study. These results were included in the meta-analyses.

[j](1) Personalized Real-Time Intervention for Stabilizing Mood (PRISM): Augmentative mobile ecological momentary intervention targeting self-management of mood symptoms. (2) Paper-and-pencil monitoring.

[k]There was a significant association between PRISM Compliance and change in YMRS Scores at 12 weeks with greater compliance associated with reduction in YMRS Score ($r = -0.453$, $p = 0.004$).

[l]Illness Intrusiveness Scale (13-item self-report scale that assesses the degree of interference caused by an illness or its treatment as a general measure of the negative functional impact of chronic illness).

[m](1) Active group: Smartphone-delivered intervention (FOCUS). (2) Control group: Clinic-based group intervention (Wellness Recovery Action Plan [WRAP]).

[n]*Effect size and p-value were not reported for the BD population only. However, the study by Ben-Zeev 2018*[47] *reported that treatment groups did not differ in change from baseline to three months post-intervention on primary and secondary clinical outcomes, including the relevant clinical outcomes assessed in this meta-analysis: BDI-II for depression, RAS for functioning and quality of life evaluations.*

[o]*The studies by Ben-Zeev 2018*[47] *and Ben-Zeev 2021*[54] *did not report specific outcomes on function. The study by Ben-Zeev 2018*[47] *reported the Recovery Assessment Scale (RAS), used to assess recovery, evaluated many aspects regarding functioning. Therefore, we reported these outcomes in the "function outcomes" section. The study by Ben-Zeev 2021*[54] reported the RAS scale and the *Sheehan Disability Scale (SDS), the latter used to assess functional disability. Both outcomes were reported in the "function outcomes" section.*

[p]The Biological Rhythms Interview of Assessment in Neuropsychiatry (BRIAN) is a 21-item evaluation considering five main areas of daily rhythms (i.e., sleep, activities, social rhythms, eating patterns, and predominant circadian rhythms).

[q]Fully remote randomized controlled crossover waitlist trial design. (1) Active group: Smartphone-delivered intervention (COREapp). Participants completed the baseline assessment (T1) and were then immediately given access to the CORE app for 30 days of use. After a month, they concluded the intervention, uninstalled the app, and completed a second assessment (T1). After an additional month (at 60 days), they completed a third assessment (T3) to measure the stability of symptom change post-intervention. (2) Control group: participants completed a baseline assessment (T1) and waited 30 days to receive the CORE app. After a month, they completed a second assessment (T2) and were provided access to the CORE app. After an additional month (at 60 days), they completed a third assessment (T3) to measure within-subject changes.

[r]The Life Goals app includes a total of 13 modules that provide self-management activities for managing the everyday needs of individuals with BD, including mood symptom coping strategies, stigma concerns, emotional self-awareness and family support, anger and irritability, and preparation for doctor's visits. The content and wording were adapted from the original Life Goals Collaborative Care (LGCC) program.

[s]All scales were self-reported. Pre-post comparisons reported were baseline to 6 months. The PHQ-9 score was used as a measure of depressive symptoms. The activation score of the ISS scale was used as an indirect measure of (hypo)manic symptoms. WHO-DAS 2.0 scale was used as an indirect measure of functioning.

As UEI, we extracted information on usability, satisfaction, acceptability, or feasibility of the app, following the model of a recently published review of user engagement in mental health apps.[58] We included the definitions used by each study of each of the UEI terms mentioned, and the outcomes of each one as defined per studies (Table 5), even if the outcomes were the same but defined differently among the studies. We included among the UEI the term "adherence" as it was widely used among the studies as an UEI, and "other UEI" for any other relevant UEI used, since some terms used to define UEI did not fit the aforementioned terms, such as "retention, dropouts or fidelity." Both objective and subjective evaluations, if present, of each UEI term were collected and reported.

"Adherence" was defined in RCTs as (i) "adherence to self-monitoring"[42–44] and objectively measured as daily completed self-monitoring,[43,44] (ii) "compliance" and objectively measured as "number of days completing a survey or an entry into a mood chart,"[45]

TABLE 4 Meta-analyses of randomized controlled trials

	k	Hedges' g[a]	95% CI	Z-value	p-value[b]	I^2
Depression						
All samples	5	−0.19	(−0.72, 0.34)	−0.69	1	85%
HDRS-17 ≥ 7	3	0.40	(0.09, 0.70)	2.56	0.074	0%
Mania						
All samples	4	−0.05	(−0.36, 0.27)	−0.31	1	56%
YMRS ≥ 7	3	0.05	(−2.23, 2.33)	0.04	1	92%
Function	4	−0.08	(−0.29, 0.12)	−0.78	1	0%
Quality of life	3	−0.12	(−0.41, 0.18)	−0.79	1	35%
Stress	3	−0.15	(−0.68, 0.38)	−0.56	1	79%

[a]Negative results favor smartphone-based interventions.

[b]Bonferroni-corrected for multiple testing.

Abbreviations: HDRS-17: Hamilton Depression Rating Scale with 17 items version; YMRS: Young Mania Rating Scale.

(iii) "mobile-device interactions mean adherence" and objectively measured as the "% of surveys responded during the monitoring period,"[46] and was not defined per se but referred as "engagement" and considered positive if "participants used the app on at least five of seven days a week"[47] or if participants completed interventions.[54] Outcomes of "adherence" ranged from 26%[47] to 93%.[44] In the observational studies, "adherence" was not precisely defined, but it was included with the term "feasibility"[48] or mixed with the terms "use" and "engagement."[49,50] In all observational studies, it was measured as "patients actively using the app at the end of the study" and outcomes ranged from 22.8%[49] to 74%.[48] Other concepts used to define or measure "adherence" were "retention" and "dropouts" regarding patients that completed the study while using the app.[48,49]

"Usability" was not defined in any of the RCTs. However, some of them reported positive feedback on the use of their smartphone-based intervention referred, such as "a useful intervention to address moderate to severe depressive symptoms,"[47] "easy to use,"[54] "user-friendly with a high usability,"[44] "usable and useful,"[42] or "acceptable to use."[43] One study derived a self-report usability and acceptability measure comprising adapted items from the SUS, Post Study System Usability Questionnaire, Technology Assessment Model Measurement Scales, and Usefulnes, Satisfaction, and Ease questionnaire.[54] In the observational studies, "usability" was defined as (i) "utility of the app according to the patient's condition and clinical state" and measured with a 5-item Likert-scale self-questionnaire,[48] not specifically defined, but measured with the System Usability Scale (SUS)[59,49] or measured with a 7-item questionnaire to assess usability and acceptability. All observational studies reported high usability measures, ranging from a mean SUS value of 77.23 of 100, lower among non-completers[49] to 82% of participants reporting that the app was useful,[48] or one study reporting that "60% of users found the app useful."[50]

"Satisfaction" was not defined in any of the RCTs, but was measured as (i) self-report ratings,[47] (ii) self-reported satisfaction scales,[45] agreement statements rating,[54] (iii) the Verona Satisfaction Scale-Affective Disorder (VSS-A),[42] and not defined nor measured in the other RCTs.[43,44,46] Outcomes of satisfaction ranged from a mean value of 25.7 of 35[47] to a median value of 9/10.[45] In the observational studies, "satisfaction" was defined as "overall experience satisfaction" and measured with a 5-item Liker-scale self-questionnaire[48,49] in two observational studies. Both studies reported high satisfaction of participants using the app, ranging from 62%[49] to 86%.[48] Another study measured it using a 7-item questionnaire and reported "poor outcomes" regarding satisfaction and perceived usefulness of the app: "a minority of the participants felt that the app helped them to make progress on their wellness goals and improved their ability to manage their own health".

"Acceptability" was not defined in any of the RCTs, but was measured as (i) a subjective evaluation from several statements, in a previous study,[47,57] (ii) a self-reported questionnaire of specific acceptability of mobile devices,[45] and not specifically measured in the other RCTs. However, some of them reported positive feedback on the acceptability of their smartphone-based intervention referred, such as "the patients expressed that the self-monitoring system was supportive, useful, quick and easy to use with a low level of intrusiveness,"[44] "acceptable to use,"[43,54] "the intervention was acceptable and usable"[47,57] or positive ratings such as "I would use this device again in the future."[45] In the observational studies, "acceptability" was defined as "discretion, lack of invasiveness, comfort with the app's daily usage, and technical difficulties experienced" and measured with a 5-item Liker-scale self-questionnaire.[48] Two other studies did not define the concept, which was referred to when describing "usability" and "satisfaction" outcomes in one study[49] and measured with a 7-item questionnaire in another.[50] All studies reported high acceptability of participants using the app, ranging from 70%[49] to 92% agreeing that the app was discrete, non-invasive, and comfortable.[48]

"Feasibility" was defined as (i) an amalgam of objective parameters of the app use[47] and (ii) an unspecific term combining satisfaction and adherence.[45] There were not clearly defined measures for "feasibility." However, many studies reported positive feedback on the feasibility of use of their smartphone-based intervention referred, such as "a feasible intervention,"[47] "feasible and acceptable"[45] and "a single intervention augmented by mobile

TABLE 5 User Engagement Indicators (UEI)

Author, year, location	Completers (*N*; %)[a]	Adherence definition (measure)	Adherence outcome	Usability definition (measure)	Usability outcome	Satisfaction definition (measure)	Satisfaction outcome
Randomized control trials							
Faurholt -Jepsen et al.; 2020, Denmark[42]	(1) 41; 87.2%[b] (2) 46; 90.2%[b]	*"Adherence to self-monitoring"* Not specifically evaluated in this study, but assessed for all the RCT studies using the Monsenso system	*"The Monsenso system has shown a high self-assessment adherence (73–97%)."*	NR	*"The Monsenso system has been shown usable and useful by patients with bipolar disorder"*	Self-assessed satisfaction with care with The Verona Satisfaction Scale-Affective Disorder (VSS-A)	There were no statistically significant differences in satisfaction with care between groups
Faurholt -Jepsen et al.; 2019, Denmark[44]	(1) 85; 87.6%[b] (2) 44; 89.8%[b]	*"Adherence to self-monitoring"* Days having completed self-monitoring	(1) 72.6% (196 of 273 days – 9 months)[b] No subjective evaluation	NR	*"The Monsenso system is easy to use and user-friendly with a high usability"*	NR	NR
Faurholt-Jepsen et al.; 2015, Denmark[43]	(1) 33; 84.6%[b] (2) 34; 87.2%[b]	*"Adherence to self-monitoring"* Daily evaluation of the subjective items in the MONARCA system	(1) 93%[b] "High level of adherence"	NR	"*Patients using the MONARCA system found it acceptable to use"*	NR	NR
Depp et al.; 2019; US[46]	(1) 77; 91%[c,k] (2) 69; 81%[c,k]	*"Mobile-device interactions, mean adherence"* % of surveys responded during the monitoring period	(1) 68,7%[c] (2) 66,2%[c]	NR	NR	NR	NR
Depp et al.; 2015; US[45]	(1) 41; 80.3%[e] (2) 41; 77.4%[e]	Referred also by study as "*compliance*". Number of days completing a phone survey (at least one of two sent) or an entry into a mood chart	(1) 65% (range 12%–97%)[e] No subjective evaluation	NR	NR	Self-reported 10-point satisfaction scales	(1) 9/10 median[e] (2) 8/10 median[e]

Acceptability definition (measure)	Acceptability outcome	Feasibility definition (measure)	Feasibility outcome	Other UEI definition (measure)	Other UEI outcome	Overall UEI evaluation
NR	NR	NR	NR	*"During recent years, there has been an increasing interest in **patient-reported outcome measures (PROMS)** as valid indicators of effect".*	NR	–
NR	*"The patients expressed that the self-monitoring system was supportive, useful, quick, and easy to use with a low level of intrusiveness."*	NR	NR	**Confidentiality** (subjective statements)	*"None of the patients expressed that they felt watched, and none were uncomfortable with having the objective smartphone data collected but saw it as a safety net."*	–
NR	*"Patients using the MONARCA system found it acceptable to use"*	NR	NR	–	–	*"The high level of adherence is believed to reflect the high usability and low level of intrusiveness from the MONARCA system, a factor contributing to a high motivation from the patients."*
NR	NR	NR	*"These results indicated that single intervention augmented by mobile intervention was feasible"*	**Fidelity**: Adapted fidelity rating scale[d]	(1) 16,0[c] (2) 15,9[c]	"*Fidelity rating scales indicated a high level of fidelity to both the SM and CBT2go"*
5-point Self-reported questionnaire of specific acceptability of mobile devices	Positive ratings described such as *" I would use this device again in the future"*	Unspecific term combining satisfaction and adherence	*"Automated mobile-phone intervention is feasible and acceptable"*	–	–	–

(Continues)

TABLE 5 (Continued)

Author, year, location	Completers (*N*; %)[a]	Adherence definition (measure)	Adherence outcome	Usability definition (measure)	Usability outcome	Satisfaction definition (measure)	Satisfaction outcome
Ben-Zeev et al.; 2018, USA[47]	(1) 75; 91.5%[f,k] (2) 74; 91,4%[f,k]	Defined as "***engagement***": participants were considered as commencing treatment if they used FOCUS or attended one WRAP session[e]. Engagement was considered: (i) used the app on at least five of seven days a week (approximately 70%) (ii) if they attended at least 60 minutes of the scheduled 90-minute group session (approximately 70%)	The groups did not differ in the proportions of participants fully engaging in all 12 weeks of intervention: (1) 26%[f] (2) 22%[f]	Not defined in the original study[47]	*"FOCUS might be a useful intervention to address moderate to severe depressive symptoms among individuals with an array of mental illnesses."*	Five self-report items post-intervention with a 7-point rating scale (mean ratings)[g]	*Satisfaction ratings were similar be- tween groups: (1): 25.7±3.8 (2): 25.5±3.6 (t = −.31, df=1, p = 0.76)*[f] *"Satisfaction with treatment did not differ across groups. Participants provided high satisfaction ratings for FOCUS and WRAP, and participants in each approach reported that it was enjoyable and interactive and helped them feel better."*
Ben-Zeev et al.; 2021, US[54]	(1) 51; 33.1%[j,k] (2) 65; 40.4%[j,k]	NR	50.6% participants completed all 53 intervention levels, whereas participants, on average, completed 35 levels (66%).	Participants were asked to rate their agreement with a series of statements about the intervention consisting of a 26-item self-report usability and acceptability measure comprising adapted items from the System Usability Scale, Post Study System Usability Questionnaire, Technology Assessment Model Measurement Scales, and Usefulness, Satisfaction, and Ease questionnaire.	Most participants reported; that the app was easy to use and sufficiently interactive; and that they did not need technical support to use CORE.	See Usability definition and measure	Participants rated that they were satisfied with the CORE intervention and that they would recommend it to a friend.
Observational studies							
Hidalgo-Mazzei et al.; 2018, Spain[49]	68; 33.8%	Defined as "**retention**" and mixed with the terms "**use**" and "**engagement**". Patients actively using the app at the end of the study (6 months)	68; 33.8%	Not defined, but measured with: System Usability Scale (SUS)[59]	Mean=77.23 (SD = 16.7), significantly lower among non-completers "*Positive outcomes regarding satisfaction and usability were mainly found among completers*"	Satisfaction and perceived helpfulness Likert-scales self-questionnaires	62% reported that they were either very satisfied or satisfied with the general experience of using the app.

Acceptability definition (measure)	Acceptability outcome	Feasibility definition (measure)	Feasibility outcome	Other UEI definition (measure)	Other UEI outcome	Overall UEI evaluation
Subjective evaluation from several statements[h,i]	*"the majority of participants found the FOCUS intervention acceptable and usable."*[h,i]	Unclearly defined as an amalgam of parameters (use of smartphone functions, N of days of use, N of interactions with the system, % of spontaneous initiation of interactions vs. prompted-use).[h]	*"FOCUS smartphone intervention is feasible among people with schizophrenia."*[h]	–	–	*"The mHealth intervention showed superior patient engagement and produced patient satisfaction and clinical and recovery outcomes that were comparable to those from a widely used clinic-based group intervention for illness management."*
See Usability definition and measure	Participants rated CORE as highly usable and acceptable. Most participants reported; that they would like to use CORE more often; that if they had access to CORE, they would use it.	NR	NR	–	–	–
Not defined (Referred to when describing "usability" and "satisfaction" outcomes)	70% strongly agreed or agreed that the app was easy to use, and its daily use was discrete and quick.	Not specifically defined. Included among the definitions of engagement, retention, and use.	*"The results of this study represent the first attempt to evaluate the feasibility of offering a large-scale wide-reaching smartphone-based psychoeducation program for BD."* However, no specific outcomes on "feasibility" are reported.	**Engagement**: calculated based on the weekly percentage of completed tasks (i.e., answering daily and weekly tests) and reading the daily psychoeducational messages. (detailed in outcome)	Average daily interaction of users while using the app: 1.8 times/day (SD=3.2) Average weekly engagement with the app of completers: 50 (25%) participants completed all tasks, 30 (15%) completed half, and 89 (44%) completed less than 25% of the tasks required.	*"We found positive outcomes regarding satisfaction and usability predominantly among completers, as well as perceived helpfulness and reported benefits to well-being and general health of all participants."*

TABLE 5 (Continued)

Author, year, location	Completers (N; %)[a]	Adherence definition (measure)	Adherence outcome	Usability definition (measure)	Usability outcome	Satisfaction definition (measure)	Satisfaction outcome
Hidalgo-Mazzei et al.; 2016, Spain[48]	36; 74%	Adherence was not specifically defined, but included with the term "feasibility". Patients actively using the app at the end of the study (3 months)	36; 74%	Utility of the app according to their condition and clinical state (5-item Likert-scale self-questionnaire)	82% report that the app was useful	Overall experience satisfaction (5-item Likert-scale self-questionnaire)	86% report satisfaction using the app
Ryan et al.; 2021, US[50]	18; 64%	Adherence was not specifically defined, but included with the term "engagement". (i) Patients actively using the app at the end of the study (6 months) (ii) N of times of app use/average of minutes of app use (iii) Modules completion	(i) 18; 64% (ii) median of 25 times /average of 154 minutes (iii) 3 (17%) participants used all modules in the app.	Not specifically defined, but measured with a 7-item Questionnaire[l]	60% found the material provided by the Life Goals app useful.	Not specifically defined, but measured with a 7-item Questionnaire[l]	40% of users reported finding the app helpful in managing their health. 36% reported finding it helpful in making progress on their wellness goals.

Abbreviations: NR, Not Reported.

[a]The rate of completers was the difference between the N of patients that completed the study and the initially allocated N of patients.

[b](1) MONARCA: daily electronic self-monitoring using smartphones including a clinical feedback loop. (2) Control group.

[c](1) CBT2go *(active intervention + SM)*; (2) Self-Monitoring Only (SM).

[d]8-item fidelity rating form adapted from the Cognitive Therapy Rating Scale for Psychosis with a score range of 0–16 administered by an unmasked clinician. Sessions were audiotaped and the therapist completed a fidelity rating each session.

[e](1) Personalized Real-Time Intervention for Stabilizing Mood (PRISM). (2) Paper-and-pencil monitoring.

[f](1) Active group: Smartphone-delivered intervention (FOCUS). (2) Control group: Clinic-based group intervention (Wellness Recovery Action Plan [WRAP]).

[g]Satisfaction was measured as the sum of five self-report items completed during the 3-month, post-intervention assessment. Participants rated the following statements with a 7-point rating scale (1, strongly disagree, to 7, strongly agree): I am satisfied with the treatment program, the treatment program helped me feel better, the treatment program was not interactive enough (reverse scored), I enjoyed the treatment program, and I would recommend the treatment program to a friend.

[h]*This information is not from the original study assessing the efficacy of the smartphone intervention in people with BD,*[47] *but from a previous study that evaluates the feasibility, acceptability, and preliminary efficacy of the FOCUS smartphone intervention for schizophrenia.*[57]

[i]Participants evaluated the acceptability and usability of the FOCUS intervention by selecting "disagree, neutral, or agree" to a series of statements including the following: "I think that I would like to use FOCUS often, I found FOCUS to be very complicated, I thought FOCUS was easy to use, I think that I would need the support of a technical person to be able to use FOCUS".

[j]Fully remote randomized controlled crossover waitlist trial design. (1) Active group: Smartphone-delivered intervention (COREapp). (2) Control group: participants completed a baseline assessment and waited 30 days to receive the CORE app.

[k]*Data not specific to the population with BD, but to the whole sample including patients with other diagnoses (BD, SCZ, and SZA in Depp 2019*[46]*; BD, MDD, SCZ, and SZA in Ben-Zeev 2018*[47] *and Ben-Zeev 2021*[54]*).*

[l]A post-study Questionnaire about the user's experiences with the app was used to assess usability and acceptability. The survey contained seven statements that participants rated their agreement with on a Likert scale, ranging from "Strongly Agree" to "Strongly Disagree." The majority of participants experienced no difficulties using the app and felt that the material was displayed adequately; however, the results showed low success in the app encouraging self-management of their own health. Only a minority of the participants felt that the app helped them to make progress on their wellness goals and improved their ability to manage their own health.

Acceptability definition (measure)	Acceptability outcome	Feasibility definition (measure)	Feasibility outcome	Other UEI definition (measure)	Other UEI outcome	Overall UEI evaluation
Discretion, lack of invasiveness, comfort with the ap's daily usage, and technical difficulties experienced (5-item Likert-scale self-questionnaire)	92% agreed that the app was discrete, non-invasive, and comfortable. Only 2% had technical difficulties	The definition of "feasibility" included adherence and app usage objective parameters stored at the cloud server (detailed in outcome)	Mean of days interacted with app: 77/90 days (SD=26,2) Interaction rate (times/day): 1.3 Daily tests completed during app use: 88% Weekly tests completed during app use: 100% "The results confirm that this particular intervention is feasible and represent a satisfactory and acceptable instrument for self-management of BD"	–	–	*"The results confirm that this particular intervention is feasible and represent a satisfactory and acceptable instrument for self-management of BD as an add-on to the usual treatment."*
Not specifically defined, but measured with a 7-item Questionnaire[l]	72% found the app easy to use. Few participants reported problems accessing the app.	Comprises the concepts of acceptability and usability, measured through the 7-item Questionnaire[l]	*"Life Goals app is feasible and acceptable for individuals with BD".*	–	–	–

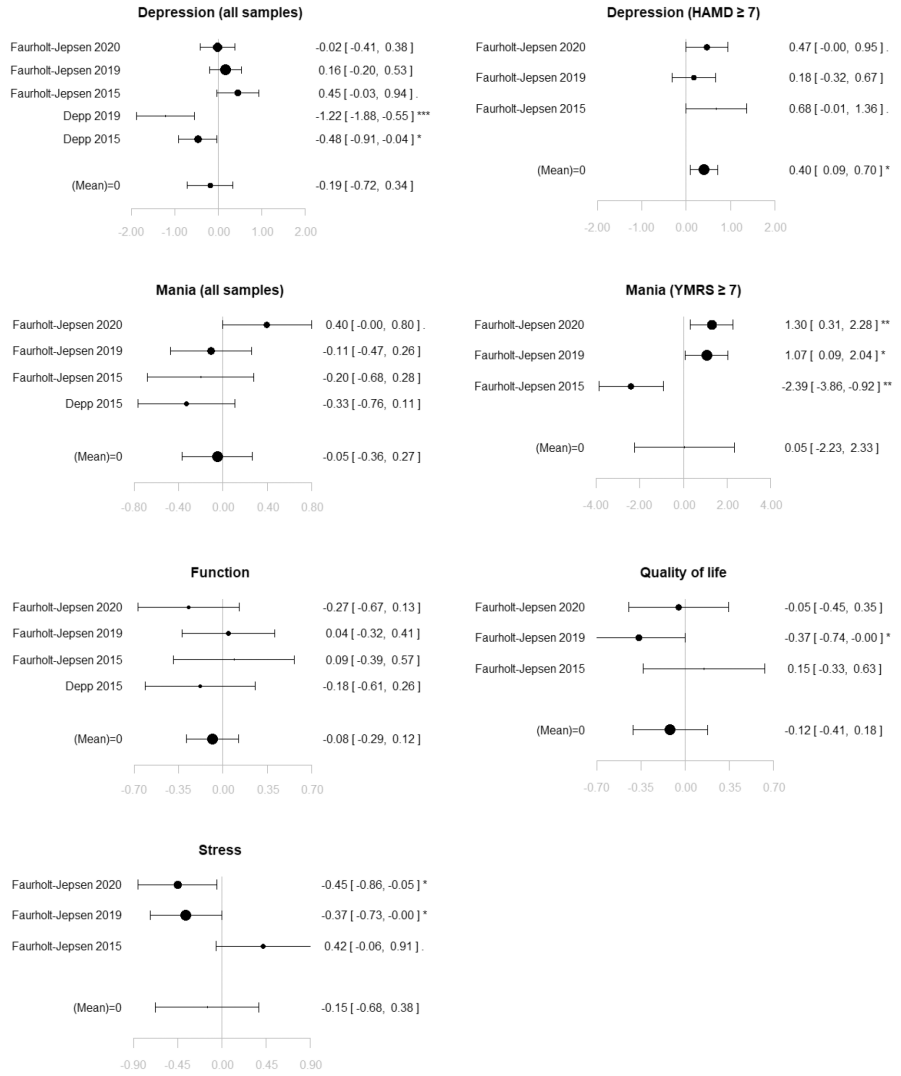

FIGURE 1 Forest plot from random-effects meta-analyses of Randomized Controlled Trials. The pre-post change scores between smartphone intervention and control conditions have been pooled to calculate the standardized difference (Hedges' g with 95% confidence intervals). Negative results favor smartphone-based interventions. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. These p-values are not Bonferroni-corrected for multiple testing (please check Table 4 for the multiple testing Bonferroni correction). Abbreviations: HAMD, Hamilton Depression Rating Scale with 17 items version; YMRS, Young Mania Rating Scale.

intervention was feasible".[46] The other RCTs did not define or refer to outcomes related to "feasibility".[42–44,54] In the observational studies, "feasibility" was defined as (i) "adherence" and measured as objective parameters of the app use,[48] (ii) an unspecific term combining the terms "engagement," "retention" and "use,"[49] and (iii) a combination of usability, acceptability and satisfaction.[50] Based on positive feedback from app use parameters, the first study concluded that their smartphone-based intervention "is feasible and represent a satisfactory and acceptable instrument for self-management of BD,"[48] the second study concluded that "positive outcomes regarding satisfaction and usability" were found, however no specific outcomes on "feasibility" were reported,[49] and the third concluded that "the app was feasible and acceptable for individuals with BD".[50]

Other UEI that were evaluated included the concepts of "confidentiality," evaluated positively with participant's statements in one study,[44] "patient-reported outcome measures (PROMS)," which have been posed as valid indicators of effect,[42] "fidelity," with positive evaluations in the active groups of one study,[46] and "engagement," which was based on the weekly percentage of completed tasks with the app.[49]

4 | DISCUSSION

This is the first comprehensive review and meta-analysis to assess the efficacy and effectiveness of smartphone-based interventions in BD considering smartphone interventions' and patient's characteristics as well as engagement measures. Our results provide the first overall estimate of the effects from such interventions and inform treatment choices and future research in this area.

A recent meta-analysis compared the effect of smartphone-based interventions and monitoring with control methods in BD.[60] This study concluded that smartphone-based interventions in BD are effective in reducing manic and depressive symptom severity. However, the analyses included studies that assessed the efficacy of phone calls of therapists to facilitate psychotherapy,[61,62] web-based platforms,[63] as well as the effectiveness of phone/mail-delivered self-rating feedback.[64] Moreover, the meta-analyses included participants with other than BD diagnoses.[46,47] Therefore, the conclusions drawn by this study may be bound to bias.

In our systematic search, most studies involving the use of smartphones in people with BD did not assess efficacy or effectiveness, but focused on the assessment of UEIs[65,66] or the correlation

TABLE 6 ISBD Big Data Task Force recommendations for consensus on studies assessing smartphone-based interventions for BD

Design		The design of the studies to evaluate efficacy or effectiveness of smartphone-based interventions in BD should probably differ from the traditional RCTs aimed at assessing the efficacy of drugs and psychological therapies.
Outcomes evaluated		As primary outcomes for efficacy and effectiveness, the risk of **affective relapses** and **psychiatric hospitalizations**, as well as the quantification of **mood instability**,[53] seem more appropriate rather than affective symptoms reduction. This might also lead **cost-effectiveness analyses** in comparison to usual care. Research groups may prefer to include one or the other measure as the primary outcome measure depending of the main objective and design of the study. Secondary outcomes may include functioning, quality of life, medication adherence, or other relevant variables. Self-reported symptoms evaluations may be bound to bias and assessed with caution. User-engagement indicators (UEIs) may influence on primary outcome results and should be considered on analyses (e.g., users with null engagement could be considered as part of the inactive control group; or users with low engagement should be sub-analyzed from users with moderate-to-high engagement). In this line, we propose the following concepts for smartphone interventions: (i) **Dose-related effect**: whether the efficacy/effectiveness of the intervention is influenced by the **use** of the app. (ii) **Clinically significant dose**: if the app is based on a clinical intervention with tested efficacy requiring a "minimum required dose" (e.g., 6 face-to-face sessions[50]), studies may consider a "minimal use of the app" related to the clinically tested efficacy (e.g., performing 6 intervention modules, using it for at least 2 months, etc.) to consider that users received a "clinically significant dose" of the intervention. Participants not using an app may be equivalent to non-adherent participants in a clinical trial and thus bias the assessments of efficacy. The proposed concepts have not been explored in the current literature and need to be considered in future studies testing smartphone interventions.
Population	**Diagnoses**	The population included in the studies should be clearly defined, including diagnoses and specifically the type of bipolar disorder. Transdiagnostic interventions may be of use. However, there are particularly specific evaluations and interventions regarding affective symptoms and psychoeducation in bipolar disorder. In studies including participants with different diagnoses, sub-analyses only on participants with BD are recommended. This way, interventions can be quantitatively compared among studies without need of requesting data to authors. Self-reported diagnoses may be bound to bias and assessed with caution.
	Severity of baseline symptoms	The baseline clinical severity of affective symptoms of participants included in the studies should be reported. Sub-analyses on participants with higher baseline depressive or manic symptoms are recommended to assess differences of the smartphone-based intervention.
Duration of the studies		The duration of the studies should be of **at least 6 months**, unless for feasibility studies. The rarity of affective episodes or psychiatric hospitalizations claims for adequate sample sizes as well as long assessment periods to identify sufficient within-subject variance.
Daily self-monitoring		The collection of daily self-monitoring of affective symptoms is recommended to: (i) provide continuous information on the psychopathological status (ii) assess mood instability,53 which may be more appropriate rather than affective symptoms reduction. (iii) combine dimensional and categorical self- and expert ratings on different time scales as latent psychopathological affective outcome variables.24
Passive data		Passive smartphone-collected data, including voice features and automatically generated objective data (calls, messages, activity, etc.) should be included in the studies to determine: (i) which combination of digital parameters can be integrated as digital phenotypes. (ii) which digital phenotypes best predict psychopathology. (iii) which combination of digital phenotypes at which days prior to an upcoming episode are viable as digital prodromal predictors. The analyses of passive smartphone-collected data should avoid fragmentation of parameters by using integrative analysis strategies such as structural equation modelling[24]

(Continues)

TABLE 6 (Continued)

Smartphone-based interventions		Smartphone-based interventions should be objectively and uniformly categorized and reported. The following are considered of paramount interest:
	General characteristics	(i) type of phone (personal or loaned). (ii) economic compensation (yes/no) and how much.
	App characteristics	(i) **psychological active therapeutic content** independent of self-monitoring and which (CBT modules, psychoeducation, cognitive restructuring, etc.). It is fundamental it is reported whether those included "active components" are **evidence-based**, and also their quality, quantity and frequency. (ii) self-monitoring (yes/no) and type (affective symptoms, adherence, stress, etc.) (iii) number of evaluations/session per a day. (iv) number of items evaluated per session (specific symptoms, type of psychological intervention, etc.) (v) collection of **passive data** (yes/no), which (activity, phone usage, etc.), when (24h, sporadically, etc.), and how (is the app connected to a **wearable** device which provides vital signs?). (vi) feedback (from clinician, from app/chatbot, clinically-oriented) and type (phone call, graphical, etc.) (vii) notifications (type, personalized, based on evaluations/interventions, etc.) (viii) chatbot (yes/no) and specific functionalities (e.g., active evidence-base component such as cognitive restructuring). (ix) natural language processing tools and which Studies assessing the efficacy of phone calls,[61,62] web-based platforms,[63] or phone/mail-delivered feedback[64] should not be considered smartphone-based interventions. Future consensus should agree on standardized measures for apps to validate their clinical application and reliability. In the latest years, platforms and databases of mHeath apps including aspects such as evidence-based validation of apps have been developed.[73,74] These may be a good start to provide future guidelines to help professionals and users.
Control groups		The characteristics of the control group should be clearly defined regarding the magnitude of intervention in this group:
	Controls using smartphone	(i) **smartphone inactive**: use of a smartphone for communicative purposes[42–44] (ii) **smartphone active**: smartphone self-monitoring without active intervention
	Controls not using smartphone (clinic-based interventions)	(i) **inactive controls**: treatment as usual (TAU).[46] (ii) **partially active controls**: paper-and-pencil monitoring[45] or phone calls from therapists.[61,62] (iii) **active controls**: clinic-based interventions.[47]
	Mixed controls: combining smartphone + clinic-based interventions	(i) **double active**: smartphone self-monitoring + clinic based controls (e.g., smartphone self-monitoring + clinic-based intervention[46]).
	Waitlist control group	The control group waits an established period of time till receiving the smartphone intervention[54]
Big data in smartphones and wearables in BD		Smartphone apps for BD can be connected to wearable devices that capture a huge amount of body signals as passive data. Smartphones and wearables in BD have shown the capacity to unobtrusively collect a huge amount of real-time objective data, which yet does not reach the level to qualify as big data. Clinicians already suffer from an overload of data. Future works need to ensure that the data incorporated provide clinically meaningful outcomes. New tools to help interpret and simplify data for use in everyday clinical decisions will be vitally important.[15] In the coming years, curators will help manage the digital data from the patient by creating a more friendly view for clinicians. The combined analysis of digital data in addition to real-time physiologic and biological information, such as cortisol, melatonin, and heart rate, seems promising for future research to provide greater insight into the physiopathology of BD. A potential limitation of the success of these approaches will be the physician's and healthcare system's ability to integrate these data into clinical practice in a way that is ethically sound, legally permissible and respectful of patients´ privacy.
Ethics and Confidentiality		Confidentiality issues are and will arise regarding the collection of digital data. New ethics protocols are needed to adapt to a constantly changing scenario.
Consumer's guide		The field of digital health needs a set of standards for quality that will include measures of efficacy, engagement, and privacy, such as with development of a consumer's guide for digital mental health, complete with user reviews.

TABLE 6 (Continued)

	UEI need consistent and replicable standard definitions and valid measures with solid thresholds. The ISBD Big Data Task force proposes that studies assessing smartphone-based interventions should report the following:						
				Thresholds for positive outcomes (Some depending on the duration of the study)[c]			
User engagement indicators (UEI)	Definition	Outcome	Magnitude for UEI	<6 months	6–9 months	9–12 months	>12 months
Completer's rate	Patients that completed the study **regardless of the use of the app**	% of patients that completed the study from those initially allocated	high	≥80%	≥70%	≥60%	≥50%
			moderate	51%–79%	41%–69%	31%–59%	21%–49%
Adherence (Retention, Fidelity)	Use of the app **at the end** of the study	% of patients using the app **at the end** of the study					
Compliance[d]	Quantification of app use during the study	% of days having completed self-monitoring/ surveys/ psychological interventions performed during the study % of days having entered into a mood chart/ feedback from app	low	≤50%	≤40%	≤30%	≤20%
Engagement[d]	Active use of the app **during the whole study**	% of weeks using the app on at least 5/7 days (approx. 70%) a week.					
Usability[a,b]	Perceived usefulness of the app, user-friendliness and complexity of usage of the system	System Usability Scale (SUS)[59,e]	high	≥80%			
Satisfaction[a,b]	Overall subjective experienced satisfaction with the app: comfort using it, organization, interfaces, etc.	Likert-scale self-questionnaire post-intervention (% of responders agreeing/ strongly agreeing with perceived satisfaction)	moderate	51%–79%			
Acceptability[f,a,b]	Perceived ease of use, discretion, lack of invasiveness, and comfort with the app's daily usage	Technology Acceptance Model (TAM) (Davis, 1989)[f]	low	≤50%			

(Continues)

TABLE 6 (Continued)

	UEI need consistent and replicable standard definitions and valid measures with solid thresholds. The ISBD Big Data Task force proposes that studies assessing smartphone-based interventions should report the following:						
				Thresholds for positive outcomes (Some depending on the duration of the study)[c]			
User engagement indicators (UEI)	Definition	Outcome	Magnitude for UEI	<6 months	6–9 months	9–12 months	>12 months
Feasibility	An app is feasible when it has proven moderate-to-high engagement, usability, satisfaction and acceptability	Mean value of the following scores: (Engagement + Usability + Satisfaction + Acceptability)/4					

[a]The mHealth App Usability Questionnaire (MAUQ)[104] was designed based on a number of existing questionnaires used in previous mobile app usability studies, including questionnaires focused on usability (SUS), acceptance (TAM), and satisfaction. Four versions of the MAUQ are available considering the type of app (interactive or standalone) and target user of the app (patient or provider). This questionnaire has simplified the existing literature on UEI, is adapted to mental health apps of different characteristics, and provides a web service for customization.[105] However, the MAUQ does not provide a cutoff point for considering an app "usable", and only states that "the higher the overall average, the higher the usability of the app".

[b]Most questionnaires mentioned evaluating UEIs usually include the following dimensions: ease of use, perceived usefulness, and satisfaction with the platform. Afterwards, the combination of these dimensions are conceptualized respectively into "acceptability", "usability", and "satisfaction". The term "usability" is usually broader than the others and includes most of these dimensions. The MAUQ subscales[104] consider the previous dimensions and could be used to quantify them.

[c]The thresholds for positive outcomes have been proposed according to the data from the studies included in the review.

[d]The thresholds for positive outcomes for compliance/engagement refer to each participant of the study. For instance, one participant that completed >90% of days entering a self-monitoring survey is highly compliant and another completing <10% is poorly compliant. Also one participant that used the app ≥5/7 days a week the 90% of weeks has a high engagement and another suing the app ≥5/7 days a week less than the 10% of weeks has a poor engagement. Once the compliance/engagement of each subject has been established, researchers should calculate the % of highly, moderate and poor compliant or engaged participants to obtain the overall compliance/engagement measure for the study (high, moderate or low).

[e]In the System Usability Scale (SUS), participants are asked to score 10 items with one of five responses that range from Strongly Agree to Strongly disagree. The final score is converted to 0–100.

[f]The Technology Acceptance Model (TAM) assesses acceptance of technology according to several constructs that include: previous experience, perceived affective quality, perceived usefulness, availability, perceived ease of use, attitude and intention to use. Each construct has one or more items that should be adapted to the technology that is being assessed (e.g., The TAM scale has been applied and validated for smartphones[102] and wearables.[103] Participants are asked to score items with one of five responses that range from Strongly Agree to Strongly disagree. The final score is converted to 0–100.

between smartphone self-reported and automatically collected data and clinician-assessed scales.[67–72]

Six RCTs evaluated the efficacy of smartphone interventions in BD comparing differences in depressive and/or manic symptoms assessed with clinician-administered scales between groups.[42–47] None of the RCTs showed efficacy in reducing depressive or manic symptoms or improving functioning when comparing to active controls, but only when compared to inactive controls (paper and pencil monitoring[45] or TAU[46]).

The first consideration is that comparisons in the RCTs ranged from inactive controls (normal use of smartphones[42–44]) to controls with highly active interventions (face-to-face intensive psychoeducational sessions plus smartphone self-monitoring).[46] In this regard, the highest efficacy was found when the difference between the smartphone intervention and the control comparison was most marked (e.g., CBT+SM smartphone intervention vs. TAU[46]) and lowest when differences between the smartphone intervention and the control comparison were less marked (control group involved additional interventions; e.g., smartphone-app vs. intensive group intervention[47]). Only one RCT used a waitlist control group and a fully remote randomized crossover design and showed efficacy in reducing depressive symptoms and improving function outcomes.[54] However, this study did not provide specific data on people with BD and diagnoses and evaluations were self-reported. This particular RCT design may be bound to limitations, but also offers a groundbreaking approach with benefits worth considering.

Moreover, the control groups in the included RCTs were highly heterogeneous among studies, thus precluding a uniform comparison. The same problem was found when trying to identify specific aspects of smartphone interventions and subpopulations with BD associated with the efficacy of the intervention. The heterogeneity in the design of the studies and the lack of uniform registered variables and definitions impeded those analyses. The mechanism of change in smartphone interventions is poorly understood, such as whether SM by itself may affect change or if therapeutic elements that draw from evidence-based interventions such as CBT are impactful beyond SM. In the latest years, platforms and databases of mHeath apps have been developed. These platforms include aspects such as evidence-based validation of apps, and thus

provide guidelines to help professionals and users. However, very few apps backed with scientific evidence focus on the needs of individuals with BD, especially beyond mood tracking.[73,74] Additionally, the dose-related effect of smartphone interventions is not usually considered among studies when assessing the efficacy or effectiveness of interventions (i.e., whether the efficacy/effectiveness is influenced by the use of the app). One study[50] introduces the concept of "clinically significant dose": the app is based on a clinical intervention with tested efficacy after completing 6 sessions (equivalent to 6 modules of the app). Thus, only participants that completed ≥6 modules were considered to have received a "clinically significant dose" of the intervention. The concepts of "minimum required dose" or "dose-related efficacy" have not been widely explored and need to be considered in future studies testing smartphone interventions. Participants not using the app may be equivalent to non-adherent participants in a clinical trial and thus bias the assessments of efficacy.

The second consideration is that most patients included in the RCTs were euthymic or with mild affective symptoms.[75] It is known that the magnitude of affective symptom reduction increases with the initial severity of symptoms both for depressive[76,77] and manic symptoms,[78] and may be minimal or non-existent, on average, in patients with mild or moderate symptoms. Probably, for this reason, the expected reduction in depressive and manic symptoms scales was non-existent to minimal in RCTs which included only euthymic or subsyndromal BD patients. This was the case in a post hoc analysis of one of the included RCTs, in which participants with moderate-to-severe depression did have significant reductions in depression symptoms at post-treatment compared to participants with minimal or mild depression.[52] However, this was not consistent in other RCTs, in which participants with more severe baseline depressive (HDRS ≥ 7) or manic symptoms (YMRS ≥ 7) experienced higher levels of depressive/manic symptoms compared with the control groups.[42–44]

Also, three of the included studies[42–44] are from the same group and use the same smartphone-based intervention (The Monsenso System). On the one hand, this may limit the generalization of the results. On the other hand, the analysis of these subsequential studies allow to consider the effects of a changing and updating smartphone intervention that added the possibility to capture passive data[79] and finally provided CBT apart from SM and added the possibility to receive psychoeducational content[42] (Table 2).

There are several limitations of smartphone-based interventions for BD that should be acknowledged. BD is a disease with high neurobiological underpinnings,[80] requiring interventions capable of modifying these altered mechanisms such as pharmacological treatments in order to control the fluctuating course of the disease.[81] In addition, psychological interventions in BD, always adjunct to pharmacotherapy have proven beneficial in reducing affective relapses, particularly depressive,[82–85] in improving cognition[86] and functioning,[87,88] as well as in overall illness management.[89] However, in acute bipolar episodes (mania or depression), adjunctive psychotherapies did not improve the rate of recovery when compared with pharmacotherapy alone in most studies.[90,91] In an illness with such a high biological load, psychological interventions alone are unlikely to stabilize the natural biphasic course of the illness.

Smartphone-based interventions for BD may be bound to the same limitations as psychotherapies: they may be useful to increase the patient's insight, adherence to medication,[92] and also to identify prodromal symptoms to prevent full-blown affective episodes. Thus, smartphone interventions may indirectly improve the patient's stability, but on their own they may not be able to reduce symptom intensity or achieve symptomatic remission and functional recovery or change the natural biphasic course of BD. Hence, the aforementioned RCTs may have been imperfect in design—by mirroring traditional RCTs aimed at assessing the efficacy of drugs and psychological therapies—and perhaps overestimating the effect of smartphones on the primary outcomes targeted.[93] It is also important to note that some smartphone apps have already demonstrated efficacy in some secondary outcomes which are considered by patients more important than symptom reduction, such as quality of life or perceived stress of illness.[44] Thus, smartphones could actually play a pivotal role in outcomes which matters most in the context of patient-centered healthcare which could be prioritized in future trials. On the other hand, it should be noted that such patient-reported outcome measures (PROMs) may be influenced by report bias due to the unblinded nature of these data.

Likewise, a post hoc analysis of the MONARCA I study proposed the measure of "mood instability" as an outcome of efficacy instead of affective symptom reduction.[53] This concept supports the idea that many patients with BD remain subsyndromally symptomatic during inter-episode periods.[94] Extensive evidence shows that mood stability is of core pathogenetic significance in BD.[79,95–98] Thus, a substantial proportion of patients with BD experience subsyndromal mood swings on a daily basis associated with increased perceived stress, decreased quality of life and functioning,[79,97] and increased risk of affective relapses and psychiatric hospitalizations.[95,96,98]

Increased mood instability behaves as a genetic vulnerability trait for BD as it is present in remitted patients[79] and in their unaffected relatives.[97] Accordingly, during the last decade there has been a gradual shift from a focus on affective episodes to inter-episodic mood instability.[94,99] Mood instability is currently considered as a new treatment target in BD as it appears to be a more sensitive measure of outcome in RCTs than more conventional outcomes focusing on affective symptom reduction.[94,99,100] This is because, unlike traditional clinician rating scales that are labor intensive and capture brief periods of time, mood instability can be captured daily by self-report for long time periods (up to years) with high adherence and low cost from a large sample of individuals thus having the potential to increase statistical power and sensitivity to detect effects in RCTs. Mood instability has internal validity as a real-life measure for patients and high external validity as it reflects patients' illness severity and functioning.

Furthermore, increased mood instability is directly associated with affective relapses and psychiatric hospitalizations.[53] Therefore, it may be a potential prodromal marker for subsyndromal affective

relapses and thus have a key role in early prevention of depressive and manic episodes.[100] It has been hypothesized that smartphone-based interventions may be effective to capture mood instability[53] and detect mood swings in inter-episodic phases to promote an early intervention to avoid affective relapses and psychiatric hospitalizations.[42,44] Therefore, the quantification of mood instability, as well as affective relapses and psychiatric hospitalizations, may be more sensitive outcomes to measure the efficacy of smartphone-based interventions in BD. Research groups may prefer to include one or the other measure as the primary outcome measure depending of the main objective and design of the study. The RCTs led by Faurholt-Jepsen's team have embraced this idea and assessed the risk of affective relapses[42,44] and psychiatric readmissions[42] during the studies. The results of these outcomes so far have been negative or conflicting, so that further studies are required to extend the existing evidence. There is an ongoing and relevant debate on how to best calculate mood instability which also highlights some of its potential limitations in the design of the studies. Thus, issues on how to collect data on mood from patients allocated to the control group should be considered. Also, issues on how to handle missing data and calculation methods applied should be addressed. Recently, there have been published guidelines on how to calculate and report mood instability.[101]

Finally, but of paramount importance, all studies concluded that their app reported positive outcomes evaluations for UEIs, except for one poor perceived usefulness and satisfaction.[50] However, our results reflect that most terms were usually conflated, used with high heterogeneity among studies and sometimes interchangeably within studies. Moreover, outcomes to measure these unclear terms varied from self-reported scales (mostly subjective and non-validated) to objective app use parameters with arbitrary thresholds (to define positive or negative outcomes) that were never defined before the study. As reported in a recent review,[58] inconsistencies in the UEI evaluation process cast doubt on the studies' ability to claim that their app "engagement" or "feasibility." However, the process by which this review defined each UEI (usability, satisfaction, acceptability, or feasibility of the app) was not specified (e.g., how information on each UEI was extracted from the studies: based on a consensus definition, as per each study defined each term, etc.). Also, the role of UEIs on analyses and their influence on primary outcome results is yet to be determined (e.g., if someone used less an app or discontinued its use in the intervention arm, how this subject data shall be handled?). Moreover, no expert-consensus has been reached so far to establish consistent and replicable definitions for UEI in smartphone-based interventions for BD.

The high heterogeneity found between RCTs comprises several aspects including the characteristics of smartphone interventions (some consisting of self-monitoring only, while others include CBT, psychoeducation, etc.), and the populations compared, both considering clinical diagnoses and the severity of affective symptoms at inclusion. Therefore, the evidence from this meta-analysis of several heterogeneous RCTs should be assessed with caution.

In sum, we did not find evidence to support that smartphone interventions may reduce the severity of depressive or manic symptoms in BD, or improve functioning, quality of life or perceived stress. The high heterogeneity of studies assessing smartphone-based interventions in BD supports the need for expert consensus to establish how studies assessing the efficacy and effectiveness of smartphone-based interventions for BD should be designed, including:

(i) Valid measures with solid evidence-based thresholds for positive outcomes
(ii) The definition of active versus inactive or partially active controls
(iii) The characteristics of the populations assessed, including the initial severity of symptoms
(iv) Consistent and replicable definitions for reporting and objectively measure:

a. Aspects of smartphone interventions (from psychotherapeutic content to technical characteristics).
b. UEIs and how they could impact on primary outcomes.
c. Cost-effectiveness.

The ISBD Big Data Task Force provides recommendations for consensus to reduce the heterogeneity and achieve more valid evidence in the field (Table 6).

The aforementioned recommendations on smartphone-based interventions for BD will allow clinicians to (i) compare and replicate studies and reach higher scientific rigor, (ii) qualitatively and quantitatively classify and rank smartphone-based interventions, and (iii) have accurate and reliable UEI to evaluate smartphone-based interventions in BD.

AUTHORS' CONTRIBUTIONS

GA, DHM, and MFJ conceived and designed the study. GA and DHM did the literature search and quality assessment of included studies. GA and DHM selected the studies and extracted the relevant information. GA and DHM synthesized the data. JR performed the meta-analyses. GA and DHM wrote the manuscript. MFJ, LVK, JR, AD, and EV critically reviewed the manuscript for intellectual content. All authors participated in the conceptualization and supervision of the study. All authors reviewed, edited, and approved the final version of the manuscript for publication. EV guided and supervised the overall work, and had final responsibility for the decision to submit for publication.

ACKNOWLEDGMENTS

Gerard Anmella is supported by a Rio Hortega 2021 grant (CM21/00017) from the from Spanish Ministry of Health financed by the Instituto de Salud Carlos III (ISCIII) and cofinanced by the Fondo Social Europeo Plus (FSE+). Dr. Hidalgo-Mazzei´s research is supported by a Juan Rodés JR18/00021 to DH granted by the ISCIII. Dr. Vieta and Dr. Radua thank the support of the Spanish Ministry of Science and Innovation (PI15/00283, PI18/00805, PI19/00394,

CPII19/00009) integrated into the Plan Nacional de I+D+I and co-financed by the Instituto de Salud Carlos III (ISCIII)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER); the ISCIII; the CIBER of Mental Health (CIBERSAM); the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2017 SGR 1365), the CERCA Programme, and the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357. Dr Passos' research is supported by National Council for Scientific and Technological Development (CNPq) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

CONFLICTS OF INTEREST

GA has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck/Otsuka, and Angelini with no financial or other relationship relevant to the subject of this article. DHM received CME-related honoraria from Abbott and Angelini with no financial or other relationship relevant to the subject of this article. EB has received research grants or research support from FAPESP, CNPq, and CAPES (Brazilian Government), from the Southeastern Ontario Academic Medical Association (SEAMO), from Faculty of Health Sciences, Queen's University and from a L'Oreal/UNESCO/Brazilian Academy of Sciences for Women in Science Award. She has received honorarium as consultant/advisory board member from Daiichi-Sankyo. ICP has received research support from or served as consultant, adviser or speaker for Lundbeck, EMS, Libbs, and receives authorship royalties from Springer Nature and ArtMed. LNY has been a consultant and/or has received speaker fees and/or sat on the advisory board and/or received research funding from Abbvie, Alkermes, Allergan, Canadian Network for Mood and Anxiety Treatments (CANMAT), Canadian Institutes of Health Research (CIHR), Dainippon Sumitomo Pharma, Gedeon Richter, GSK, Intracellular Therapies, Lundbeck, Merck, Otsuka, Sanofi and Sunovion over the past 3 years. RWL has received honoraria for ad hoc speaking or advising/consulting, or received research funds, from Allergan, Asia-Pacific Economic Cooperation, BC Leading Edge Foundation, Canadian Institutes of Health Research, Canadian Network for Mood and Anxiety Treatments, Healthy Minds Canada, Janssen, Lundbeck, Lundbeck Institute, Michael Smith Foundation for Health Research, MITACS, Myriad Neuroscience, Ontario Brain Institute, Otsuka, Pfizer, Unity Health, and VGH-UBCH Foundation. RSM has received research grant support from CIHR/GACD/Chinese National Natural Research Foundation; speaker/consultation fees from Lundbeck, Janssen, Purdue, Pfizer, Otsuka, Takeda, Neurocrine, Sunovion, Bausch Health, Novo Nordisk, Kris, Sanofi, Eisai, Intra-Cellular, NewBridge Pharmaceuticals, Abbvie. RM is a CEO of Braxia Scientific Corp and has received research grant support from CIHR/GACD/Chinese National Natural Research Foundation; speaker/consultation fees from Lundbeck, Janssen, Alkermes, Mitsubishi Tanabe, Purdue, Pfizer, Otsuka, Takeda, Neurocrine, Sunovion, Bausch Health, Novo Nordisk, Kris, Sanofi, Eisai, Intra-Cellular, NewBridge Pharmaceuticals, Abbvie. EV has received research support from or served as consultant, adviser, or speaker for AB-Biotics, Abbott, Allergan, Angelini, Dainippon Sumitomo Pharma, Ferrer, Gedeon Richter, Janssen, Lundbeck, Otsuka, Sage pharmaceuticals, Sanofi-Aventis, Shire, Sunovion, Takeda, and reports no financial or other relationship relevant to the subject of this article. All other authors report no financial or other relationship relevant to the subject of this article. LVK has during the recent 3 years been a consultant for Lundbeck and Teva.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author (EV) on reasonable request. The corresponding author had full access to all the data in the study.

ORCID

Gerard Anmella https://orcid.org/0000-0002-6798-4054
Maria Faurholt-Jepsen https://orcid.org/0000-0002-0462-6444
Diego Hidalgo-Mazzei https://orcid.org/0000-0002-2693-6849
Joaquim Radua https://orcid.org/0000-0003-1240-5438
Ives C. Passos https://orcid.org/0000-0001-6407-8219
Flavio Kapczinski https://orcid.org/0000-0001-8738-856X
Martin Alda https://orcid.org/0000-0001-9544-3944
Tomas Hajek https://orcid.org/0000-0003-0281-8458
Boris Birmaher https://orcid.org/0000-0001-9299-6519
Danella Hafeman https://orcid.org/0000-0001-8312-3513
Tina Goldstein https://orcid.org/0000-0002-4762-8060
Elisa Brietzke https://orcid.org/0000-0003-2697-1342
Anne Duffy https://orcid.org/0000-0002-5895-075X
Benno Haarman https://orcid.org/0000-0002-9006-8863
Carlos López-Jaramillo https://orcid.org/0000-0002-1875-1369
Lakshmi N. Yatham https://orcid.org/0000-0002-7405-0954
Raymond W. Lam https://orcid.org/0000-0001-7142-4669
Erkki Isometsa https://orcid.org/0000-0001-5956-2399
Rodrigo Mansur https://orcid.org/0000-0002-3968-3297
Roger S. McIntyre https://orcid.org/0000-0003-4733-2523
Benson Mwangi https://orcid.org/0000-0002-1717-4395
Eduard Vieta https://orcid.org/0000-0002-0548-0053
Lars Vedel Kessing https://orcid.org/0000-0001-9377-9436

REFERENCES

1. Carvalho AF, Firth J, Vieta E. Bipolar disorder. Ropper AH, ed. *N Engl J Med*. 2020;383(1):58-66. doi:10.1056/NEJMra1906193
2. Salagre E, Dodd S, Aedo A, et al. Toward precision psychiatry in bipolar disorder: staging 2.0. *Front Psych*. 2018;9:641. doi:10.3389/fpsyt.2018.00641
3. Post RM, Fleming J, Kapczinski F. Neurobiological correlates of illness progression in the recurrent affective disorders. *J Psychiatr Res*. 2012;46(5):561-573. doi:10.1016/j.jpsychires.2012.02.004
4. Staudt Hansen P, Frahm Laursen M, Grøntved S, Puggard Vogt Straszek S, Licht RW, Nielsen RE. Increasing mortality gap for patients diagnosed with bipolar disorder-A nationwide study with 20 years of follow-up. *Bipolar Disord*. 2019;21(3):270-275. doi:10.1111/bdi.12684
5. Kessing LV, Vradi E, Andersen PK. Life expectancy in bipolar disorder. *Bipolar Disord*. 2015;17(5):543-548. doi:10.1111/bdi.12296
6. Ferrari AJ, Stockings E, Khoo JP, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord*. 2016;18(5):440-450. doi:10.1111/bdi.12423

7. Ching CRK, Hibar DP, Gurholt TP, et al. What we learn about bipolar disorder from large-scale neuroimaging: findings and future directions from the ENIGMA Bipolar Disorder Working Group. *Hum Brain Mapp*. 2020;43(1):56-82. doi:10.1002/hbm.25098
8. Kishi T, Ikuta T, Matsuda Y, et al. Mood stabilizers and/or antipsychotics for bipolar disorder in the maintenance phase: a systematic review and network meta-analysis of randomized controlled trials. *Mol Psychiatry*. 2020;26(8):4146-4157. doi:10.1038/s41380-020-00946-6
9. Vieta E, Salagre E, Grande I, et al. Early intervention in bipolar disorder. *Am J Psychiatry*. 2018;175(5):411-426. doi:10.1176/appi.ajp.2017.17090972
10. Kessing LV, Vradi E, Andersen PK. Starting lithium prophylaxis early v. late in bipolar disorder. *Br J Psychiatry*. 2014;205(3):214-220. doi:10.1192/BJP.BP.113.142802
11. de Assis da Silva R, Mograbi DC, LAS S, et al. Insight across the different mood states of bipolar disorder. *Psychiatry Q*. 2015;86(3):395-405. doi:10.1007/s11126-015-9340-z
12. Pacchiarotti I, Anmella G, Colomer L, Vieta E. How to treat mania. *Acta Psychiatr Scand*. 2020;142(3):173-192. doi:10.1111/acps.13209
13. Hidalgo-Mazzei D, Young AH. Psychiatry foretold. *Aust N Z J Psychiatry*. 2019;53(4):365-366. doi:10.1177/0004867418816821
14. Hidalgo-Mazzei D, Young AH, Vieta E, Colom F. Behavioural biomarkers and mobile mental health: a new paradigm. *Int J Bipolar Disord*. 2018;6(1):9. doi:10.1186/s40345-018-0119-7
15. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol*. 2015;33(5):462-463. doi:10.1038/nbt.3223
16. Seppälä J, De Vita I, Jämsä T, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review and link to the M-REsist project. *J Med Internet Res*. 2019;6(2):e9819. doi:10.2196/mental.9819
17. Melbye S, Kessing LV, Bardram JE, Faurholt-Jepsen M. Smartphone-based self-monitoring, treatment, and automatically generated data in children, adolescents, and young adults with psychiatric disorders: systematic review. *JMIR Ment Health*. 2020;7(10):e17453. doi:10.2196/17453
18. Saccaro LF, Amatori G, Cappelli A, Mazziotti R, Dell'Osso L, Rutigliano G. Portable technologies for digital phenotyping of bipolar disorder: a systematic review. *J Affect Disord*. 2021;295:323-338. doi:10.1016/J.JAD.2021.08.052
19. Scott J, Murray G, Henry C, et al. Activation in bipolar disorders a systematic review. *JAMA Psychiat*. 2017;74(2):189-196. doi:10.1001/jamapsychiatry.2016.3459
20. Faurholt-Jepsen M, Busk J, Frost M, et al. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*. 2016;6(7):e856. doi:10.1038/tp.2016.123
21. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *J Med Internet Res*. 2019;21(4):e11029. doi:10.2196/11029
22. Stanislaus S, Vinberg M, Melbye S, et al. Smartphone-based activity measurements in patients with newly diagnosed bipolar disorder, unaffected relatives and control individuals. *Int J Bipolar Disord*. 2020;8(1):32. doi:10.1186/s40345-020-00195-0
23. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. *Int J Bipolar Disord*. 2018;6(1):6. doi:10.1186/s40345-017-0110-8
24. Ebner-Priemer UW, Mühlbauer E, Neubauer AB, et al. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. *Int J Bipolar Disord*. 2020;8(1):35. doi:10.1186/s40345-020-00210-4
25. Antosik-Wójcińska AZ, Dominiak M, Chojnacka M, et al. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. *Int J Med Inform*. 2020;138:104131. doi:10.1016/j.ijmedinf.2020.104131
26. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. *Brain Sci*. 2017;7(11). doi:10.3390/brainsci7110150
27. Insel TR. Digital phenotyping: technology for a new science of behavior. *J Am Med Assoc*. 2017;318(13):1215-1216. doi:10.1001/jama.2017.11295
28. Daus H, Kislicyn N, Heuer S, Backenstrass M. Disease management apps and technical assistance systems for bipolar disorder: investigating the patients´ point of view. *J Affect Disord*. 2018;229:351-357. doi:10.1016/j.jad.2017.12.059
29. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. *J Med Internet Res*. 2015;17(8):e198. doi:10.2196/jmir.4581
30. Lagan S, Ramakrishnan A, Lamont E, Ramakrishnan A, Frye M, Torous J. Digital health developments and drawbacks: a review and analysis of top-returned apps for bipolar disorder. *Int J Bipolar Disord*. 2020;8(1):39. doi:10.1186/s40345-020-00202-4
31. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*. 2017;16(3):287-298. doi:10.1002/wps.20472
32. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J Affect Disord*. 2017;218:15-22. doi:10.1016/J.JAD.2017.04.046
33. Goreis A, Felnhofer A, Kafka JX, Probst T, Kothgassner OD. Efficacy of self-management smartphone-based apps for post-traumatic stress disorder symptoms: a systematic review and meta-analysis. *Front Neurosci*. 2020;14:3. doi:10.3389/fnins.2020.00003
34. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
35. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak*. 2007;7. doi:10.1186/1472-6947-7-16
36. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi:10.1136/bmj.l4898
37. Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res*. 2020;7(1):1-11. doi:10.1186/s40779-020-00238-8
38. National Institutes of Health. Study Quality Assessment Tools|NHLBI, NIH - Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group. Accessed June 9, 2021. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
39. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing; 2020. Accessed July 22, 2021. https://www.r-project.org/
40. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials*. 2007;28(2):105-114. doi:10.1016/j.cct.2006.04.004
41. Albajes-Eizagirre A, Solanes A, Radua J. Meta-analysis of non-statistically significant unreported effects. *Stat Methods Med Res*. 2019;28(12):3741-3754. doi:10.1177/0962280218811349
42. Faurholt-Jepsen M, Lindbjerg Tønning M, Frost M, et al. Reducing the rate of psychiatric re-admissions in bipolar disorder using smartphones-the RADMIS trial. *Acta Psychiatr Scand*. 2021;143(5):453-465. doi:10.1111/acps.13274

43. Faurholt-Jepsen M, Frost M, Ritz C, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. *Psychol Med*. 2015;45(13):2691-2704. doi:10.1017/S0033291715000410
44. Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. *Psychol Med*. 2020;50(5):838-884. doi:10.1017/S0033291719000710
45. Depp CA, Ceglowski J, Wang VC, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. *J Affect Disord*. 2015;174:23-30. doi:10.1016/j.jad.2014.10.053
46. Depp CA, Perivoliotis D, Holden J, Dorr J, Granholm EL. Single-session mobile-augmented intervention in serious mental illness: a three-arm randomized controlled trial. *Schizophr Bull*. 2019;45(4):752-762. doi:10.1093/schbul/sby135
47. Ben-Zeev D, Brian RM, Jonathan G, et al. Mobile health (mHealth) versus clinic-based group intervention for people with serious mental illness: a randomized controlled trial. *Psychiatr Serv*. 2018;69(9):978-985. doi:10.1176/appi.ps.201800063
48. Hidalgo-Mazzei D, Mateu A, Reinares M, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. *J Affect Disord*. 2016;200:58-66. doi:10.1016/j.jad.2016.04.042
49. Hidalgo-Mazzei D, Reinares M, Mateu A, et al. OpenSIMPLe: a real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. *J Affect Disord*. 2018;241:436-445. doi:10.1016/j.jad.2018.08.048
50. Ryan KA, Smith SN, Yocum AK, et al. The life goals self-management mobile app for bipolar disorder: consumer feasibility, usability, and acceptability study. *JMIR Form Res*. 2021;5(12):e32450. doi:10.2196/32450
51. Hidalgo-Mazzei D, Reinares M, Mateu A, et al. Is a SIMPLe smartphone application capable of improving biological rhythms in bipolar disorder? *J Affect Disord*. 2017;223:10-16. doi:10.1016/j.jad.2017.07.028
52. Ben-Zeev D, Buck B, Chu PV, Razzano L, Pashka N, Hallgren KA. Transdiagnostic mobile health: smartphone intervention reduces depressive symptoms in people with mood and psychotic disorders. *JMIR Ment Health*. 2019;6(4):e13202. doi:10.2196/13202
53. Faurholt-Jepsen M, Ritz C, Frost M, et al. Mood instability in bipolar disorder type I versus type II-continuous daily electronic self-monitoring of illness activity using smartphones. *J Affect Disord*. 2015;186:342-349. doi:10.1016/j.jad.2015.06.026
54. Ben-Zeev D, Chander A, Tauscher J, et al. A smartphone intervention for people with serious mental illness: fully remote randomized controlled trial of CORE. *J Med Internet Res*. 2021;23(11):e29201. doi:10.2196/29201
55. Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. *Psychol Med*. 2020;50(5):838-848. doi:10.1017/S0033291719000710
56. Cook JA, Copeland ME, Jonikas JA, et al. Results of a randomized controlled trial of mental illness self-management using wellness recovery action planning. *Schizophr Bull*. 2012;38(4):881-891. doi:10.1093/schbul/sbr012
57. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophr Bull*. 2014;40(6):1244-1253. doi:10.1093/schbul/sbu033
58. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatr Serv*. 2019;70(7):538-544. doi:10.1176/appi.ps.201800519
59. Brooke J. SUS - a quick and dirty usability scale. *Usability Eval Ind*. 1996;189:4-7.
60. Liu J-Y, Xu K-K, Zhu G-L, Zhang Q-Q, Li X-M. Effects of smartphone-based interventions and monitoring on bipolar disorder: a systematic review and meta-analysis. *World J Psychiatry*. 2020;10(11):272-285. doi:10.5498/wjp.v10.i11.272
61. Celano CM, Gomez-Bernal F, Mastromauro CA, et al. A positive psychology intervention for patients with bipolar depression: a randomized pilot trial. *J Ment Health*. 2020;29(1):60-68. doi:10.1080/09638237.2018.1521942
62. Kilbourne AM, Prenovost KM, Liebrecht C, et al. Randomized controlled trial of a collaborative care intervention for mood disorders by a national commercial health plan. *Psychiatr Serv*. 2019;70(3):219-224. doi:10.1176/appi.ps.201800336
63. Gliddon E, Cosgrove V, Berk L, et al. A randomized controlled trial of MoodSwings 2.0: an internet-based self-management program for bipolar disorder. *Bipolar Disord*. 2019;21(1):28-39. doi:10.1111/bdi.12669
64. Miklowitz DJ, Price J, Holmes EA, et al. Facilitated integrated mood management for adults with bipolar disorder. *Bipolar Disord*. 2012;14(2):185-197. doi:10.1111/j.1399-5618.2012.00998.x
65. Van Til K, McInnis MG, Cochran A. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. *Bipolar Disord*. 2020;22(2):182-190. doi:10.1111/bdi.12849
66. Patoz MC, Hidalgo-Mazzei D, Pereira B, et al. Patients' adherence to smartphone apps in the management of bipolar disorder: a systematic review. *Int J Bipolar Disord*. 2021;9(1):1-15. doi:10.1186/S40345-021-00224-6/TABLES/3
67. Beiwinkel T, Kindermann S, Maier A, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. *JMIR Ment Health*. 2016;3(1):e2. doi:10.2196/mental.4560
68. Ryan KA, Babu P, Easter R, et al. A smartphone app to monitor mood symptoms in bipolar disorder: development and usability study. *JMIR Ment Health*. 2020;7(9):e19476. doi:10.2196/19476
69. Schärer LO, Krienke UJ, Graf SM, Meltzer K, Langosch JM. Validation of life-charts documented with the personal life-chart app - a self-monitoring tool for bipolar disorder. *BMC Psychiatry*. 2015;15(1):49. doi:10.1186/s12888-015-0414-0
70. Tsanas A, Saunders KEA, Bilderbeck AC, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *J Affect Disord*. 2016;205:225-233. doi:10.1016/j.jad.2016.06.065
71. Chan EC, Sun Y, Aitchison KJ, Sivapalan S. Mobile app-based self-report questionnaires for the assessment and monitoring of bipolar disorder: systematic review. *JMIR Form Res*. 2021;5(1):e13770. doi:10.2196/13770
72. Ortiz A, Maslej MM, Husain MI, Daskalakis ZJ, Mulsant BH. Apps and gaps in bipolar disorder: a systematic review on electronic monitoring for episode prediction. *J Affect Disord*. 2021;295:1190-1200. doi:10.1016/J.JAD.2021.08.140
73. Mobile Health Index and Navigation Database, App Evaluation Resources from the Division of Digital Psychiatry at BIDMC. Accessed February 1, 2022. https://mindapps.org/
74. One Mind PsyberGuide|A Mental Health App Guide. Accessed February 1, 2022. https://onemindpsyberguide.org/
75. Tohen M, Frank E, Bowden CL, et al. The International Society for Bipolar Disorders (ISBD) Task Force report on the nomenclature of course and outcome in bipolar disorders. *Bipolar Disord*. 2009;11(5):453-473. doi:10.1111/j.1399-5618.2009.00726.x
76. Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002;22(1):40-45. http://www.ncbi.nlm.nih.gov/pubmed/11799341
77. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *J Am Med Assoc*. 2010;303(1):47-53. doi:10.1001/jama.2009.1943

78. Samara MT, Goldberg Y, Levine SZ, et al. Initial symptom severity of bipolar I disorder and the efficacy of olanzapine: a meta-analysis of individual participant data from five placebo-controlled studies. *Lancet Psychiatry*. 2017;4(11):859-867. doi:10.1016/S2215-0366(17)30331-0
79. Faurholt-Jepsen M, Frost M, Busk J, et al. Is smartphone-based mood instability associated with stress, quality of life, and functioning in bipolar disorder? *Bipolar Disord*. 2019;21(7):611-620. doi:10.1111/BDI.12796
80. Vieta E, Berk M, Schulze TG, et al. Bipolar disorders. *Nat Rev Dis Primers*. 2018;4:18008. doi:10.1038/nrdp.2018.8
81. Carvalho AF, Solmi M, Husain MI, Berk M, Vieta E. The rebirth of lithium as the archetypal mood stabilizer. *Bipolar Disord*. 2021;23(2):211-212. doi:10.1111/bdi.13031
82. Colom F, Vieta E, Sánchez-Moreno J, et al. Group psychoeducation for stabilised bipolar disorders: 5-year outcome of a randomised clinical trial. *Br J Psychiatry*. 2009;194(3):260-265. doi:10.1192/bjp.bp.107.040485
83. Colom F, Vieta E, Martínez-Arán A, et al. A randomized trial on the efficacy of group psychoeducation in the prophylaxis of recurrences in bipolar patients whose disease is in remission. *Arch Gen Psychiatry*. 2003;60(4):402-407. doi:10.1001/archpsyc.60.4.402
84. Miklowitz DJ. Adjunctive psychotherapy for bipolar disorder: state of the evidence. *Am J Psychiatry*. 2008;165(11):1408-1419. doi:10.1176/appi.ajp.2008.08040488
85. da Costa RT, Cheniaux E, Rosaes PAL, et al. The effectiveness of cognitive behavioral group therapy in treating bipolar disorder: a randomized controlled study. *Rev Bras Psiquiatr*. 2011;33(2):144-149. doi:10.1590/s1516-44462011005000008
86. Lewandowski KE, Sperry SH, Cohen BM, et al. Treatment to enhance cognition in bipolar disorder (TREC-BD): efficacy of a randomized controlled trial of cognitive remediation versus active control. *J Clin Psychiatry*. 2017;78(9):e1242-e1249. doi:10.4088/JCP.17m11476
87. Torrent C, Bonnin CM, Martínez-Arán A, et al. Efficacy of functional remediation in bipolar disorder: a multicenter randomized controlled study. *Am J Psychiatry*. 2013;170(8):852-859. doi:10.1176/appi.ajp.2012.12070971
88. Bonnin CM, Torrent C, Arango C, et al. Functional remediation in bipolar disorder: 1-year follow-up of neurocognitive and functional outcome. *Br J Psychiatry*. 2016;208(1):87-93. doi:10.1192/bjp.bp.114.162123
89. Miklowitz DJ, Efthimiou O, Furukawa TA, et al. Adjunctive psychotherapy for bipolar disorder: a systematic review and component network meta-analysis. *JAMA Psychiat*. 2021;78(2):141-150. doi:10.1001/JAMAPSYCHIATRY.2020.2993
90. Miller IW, Solomon DA, Ryan CE, Keitner GI. Does adjunctive family therapy enhance recovery from bipolar I mood episodes? *J Affect Disord*. 2004;82(3):431-436. doi:10.1016/j.jad.2004.01.010
91. Clarkin JF, Carpenter D, Hull J, Wilner P, Glick I. Effects of psychoeducational intervention for married patients with bipolar disorder and their spouses. *Psychiatr Serv*. 1998;49(4):531-533. doi:10.1176/ps.49.4.531
92. Reinares M, Sánchez-Moreno J, Fountoulakis KN. Psychosocial interventions in bipolar disorder: what, for whom, and when. *J Affect Disord*. 2014;156:46-55. doi:10.1016/j.jad.2013.12.017
93. Eiring Ø, Nylenna M, Nytrøen K. Patient-important outcomes in the long-term treatment of bipolar disorder: a mixed-methods approach investigating relative preferences and a proposed taxonomy. *Patient*. 2015;9(2):91-102. doi:10.1007/S40271-015-0128-X
94. Bonsall MB, Wallace-Hadrill SMA, Geddes JR, Goodwin GM, Holmes EA. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *Proc R Soc B Biol Sci*. 2012;279(1730):916-924. doi:10.1098/rspb.2011.1246
95. O'Donnell LA, Ellis AJ, Van de Loo MM, et al. Mood Instability as a Predictor of Clinical and Functional Outcomes in Adolescents with Bipolar I and Bipolar II Disorder HHS Public Access. *J Affect Disord*. 2018;236:199-206. doi:10.1016/j.jad.2018.04.021
96. Patel R, Lloyd T, Jackson R, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ Open*. 2015;5(5):e007504. doi:10.1136/BMJOPEN-2014-007504
97. Stanislaus S, Faurholt-Jepsen M, Vinberg M, et al. Mood instability in patients with newly diagnosed bipolar disorder, unaffected relatives, and healthy control individuals measured daily using smartphones. *J Affect Disord*. 2020;271:336-344. doi:10.1016/J.JAD.2020.03.049
98. Strejilevich SA, Martino DJ, Murru A, et al. Mood instability and functional recovery in bipolar disorders. *Acta Psychiatr Scand*. 2013;128(3):194-202. doi:10.1111/acps.12065
99. Bopp JM, Miklowitz DJ, Goodwin GM, Stevens W, Rendell JM, Geddes JR. The longitudinal course of bipolar disorder as revealed through weekly text messaging: a feasibility study. *Bipolar Disord*. 2010;12(3):327-334. doi:10.1111/J.1399-5618.2010.00807.X
100. Broome MR, Saunders KEA, Harrison PJ, Marwaha S. Mood instability: significance, definition and measurement. *Br J Psychiatry*. 2015;207(4):283-285. doi:10.1192/BJP.BP.114.158543
101. Faurholt-Jepsen M, Geddes JR, Goodwin GM, et al. Reporting guidelines on remotely collected electronic mood data in mood disorder (eMOOD)—recommendations. *Transl Psychiatry*. 2019;9(1):162. doi:10.1038/S41398-019-0484-8
102. Joo J, Sang Y. Exploring Koreans' smartphone usage: an integrated model of the technology acceptance model and uses and gratifications theory. *Comput Hum Behav*. 2013;29(6):2512-2518. doi:10.1016/J.CHB.2013.06.002
103. Dutot V, Bhatiasevi V, Bellallahom N. Applying the technology acceptance model in a three-countries study of smartwatch adoption. *J High Technol Managem Res*. 2019;30(1):1-14. doi:10.1016/J.HITECH.2019.02.001
104. Zhou L, Bao J, Setiawan I, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. *JMIR Mhealth Uhealth*. 2019;21(4):e11500. doi:10.2196/11500
105. PITT Usability Questionnaire. Accessed February 1, 2022. https://ux.hari.pitt.edu/v2/portal/#/

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Anmella G, Faurholt-Jepsen M, Hidalgo-Mazzei D, et al. Smartphone-based interventions in bipolar disorder: Systematic review and meta-analyses of efficacy. A position paper from the International Society for Bipolar Disorders (ISBD) Big Data Task Force. *Bipolar Disord*. 2022;24:580-614. doi: 10.1111/bdi.13243