Mind the Gap(s) in Mental Health Care. On the Claimed and Actual Evidence-Base of Publicly Available Top-Rated mHealth Applications for Depression: a Scoping Review.

May., L.C.

Department of Psychology, Health & Technology, University of Twente

Student number: 1817485

First Supervisor: Youri Derks

Second Supervisor: Saskia Kelders

May 22nd, 2023

Abstract

This scoping review explores the top-rated mHealth applications for depression through an app and literature review. The objective is to examine the key features and evaluate the claimed and actual evidence base of these apps, with a specific focus on the implementation of evidence-based theoretical background/strategies. The ARI was used to score the actual evidence of the applications. The review of eleven resulting apps provides valuable insights into their characteristics and claimed evidence base. However, determining the alignment between claims and actual evidence base, especially regarding CBT implementation, remains challenging. Notably, four apps demonstrate an evidence-based theoretical background. Among them, Youper stands out as the most consistent in reflecting its claims. However, a discrepancy between developer claims or advertising and the actual evidence base is evident across the apps.

This research underscores the ongoing need to enhance the evidence base and theoretical foundations of mental health apps. It emphasizes the importance of detailed assessments, including in-app evaluations and studies, to bridge the gap between claims and evidence. Additionally, the development of valid and accessible scoring forms for professionals and non-professionals is crucial, given the limited availability of research tools.

In conclusion, this scoping review provides insights into the evidence base of top-rated mHealth apps for depression, emphasizing the necessity for continued research in the field. Future investigations should prioritize detailed examinations of evidence base and the development of user-friendly scoring tools to ensure the delivery of high-quality, evidence-based mental health interventions through mobile applications.

Keywords: depression, mHealth, application, evidence base, mobile, top rated

Table of Contents

Mind the Gap(s) in Mental Health Care. On the Claimed and Actual Evidence-Base of Publicly Available Top-Rated mHealth Applications for Depression: a Scoping Review

Introduction

If an individual were to enter "depression app" into the Google search engine on November 10th, 2022, they would be overwhelmed with over 291,000,000 search results. The Apple App Store and Google Play Store offer a great collection of applications, with recent statistics from Ceci (2022) indicating that the Google Play store currently hosts approximately 3.55 million apps, whereas the Apple App Store offers around 1.6 million apps. In a report by the IQVIA Institute for Human Data Science, it was noted that the number of consumer digital health apps increased in 2020, with over 90,000 new apps introduced (Aitken et al., 2021). The report revealed that there are now more than 350,000 digital health apps available to consumers, with 47% focusing on managing specific diseases or health conditions, compared to 28% in 2015 (Pohl, 2023). Many of these apps are self-help tools, offering skills, exercises, or information on various health conditions without the guidance of a coach or therapist (Wasil et al., 2019). The options available range from meditation and mindfulness to journaling, self-monitoring, AI chatbots, and motivational quotes.

Depression is among the most prevalent mental health disorders, affecting approximately 1 in 15 individuals each year (Kessler et al., 2005). Despite the existence of evidence-based treatment options such as Cognitive Behavioral Therapy (CBT), several barriers impede individuals with depression from receiving traditional treatment (Nice Guidelines, 2022; Mitchell & Selmes, 2007). Given that depression affects an estimated 3.8% of the global population over their lifetime (WHO, 2021), which is almost double the population of Russia, it becomes challenging to provide traditional care to all those in need (Ben-Zeev et al., 2014;

Dulin, Gonzalez, & Campbell, 2014). Studies such as the meta-analysis conducted by Linardon et al. (2019) have highlighted the potential benefits of mHealth interventions, including CBT, in addressing barriers such as cost, transportation, therapist availability, and lengthy therapy waitlists (Ben-Zeev et al., 2014; Heffner et al., 2014; Roepke et al., 2015). Consequently, further exploration of publicly available mHealth applications, particularly depression apps, is crucial to ascertain their potential as alternative or complementary options to traditional treatment approaches (Morris & Aguilera, 2012).

Despite the promising nature of mHealth and depression apps, it is imperative to examine the current evidence base or empirical support for their effectiveness in treating depression. The increasing availability of mHealth apps has not been accompanied by equal attention to their efficacy, particularly concerning publicly available apps (Lui et al., 2017). Empirical trials for mHealth apps accessible to the public are scarce (Donker et al., 2013; Lau et al., 2020). For instance, a recent systematic review of over 1,000 publicly available mHealth apps revealed that only 2% of apps had undergone scientific testing (Lau et al., 2020). Additionally, numerous articles highlight the failure of many publicly available apps to integrate evidence-based practices, a theoretical background, or clinical expertise within their design (Shen et al., 2015). It is important to note that anyone can create and release a depression app without subjecting its content to rigorous testing (Apple Inc., n.d.; Google Developers, n.d.). Clinical trials serve the purpose of evaluating the safety and efficacy of treatments, which ultimately contributes to the well-being of users utilizing depression apps (Hoffmann et al., 2017).

Depression

Depression, also referred to as major depressive disorder or clinical depression, is recognized as a prevalent and serious mood disorder according to the Diagnostic and Statistical

Manual, 5th edition (DSM-5) (American Psychiatric Association, 2022; Uher et al., 2014). The DSM-5 provides a comprehensive framework for diagnosing depression, delineating nine criteria that are indicative of the disorder. These criteria encompass the presence of five or more symptoms persisting over a span of two weeks, where at least one symptom must manifest as a depressed mood or a diminished capacity to experience pleasure or interest. In addition to the aforementioned criteria, individuals with depression may exhibit other symptoms such as alterations in appetite, disruptions in sleep patterns, fatigue, pervasive feelings of worthlessness or guilt, difficulties with concentration, and recurrent thoughts of death or suicide. (American Psychiatric Association, 2022).

While sadness is a normal emotion experienced by everyone, depression is a serious mental illness that persists and significantly impairs one's ability to function (Nock et al., 2009; Gil & Droit-Volet, 2009). Unlike transient sadness, depression requires professional attention and should be treated as a severe illness (Roth et al., 2018). Tragically, around 800,000 people die by suicide each year, making it the fourth leading cause of death among 15 to 29-year-olds. Depression and suicide are closely linked, with individuals suffering from depression being 25 times more likely to attempt suicide (Bamonti et al., 2014; Ng et al., 2017). Major depressive disorder (MDD), the more severe form of depression, is associated with a lifetime prevalence of suicide attempts estimated at 31% (Dong et al., 2019; Nock et al., 2009).

The treatment approach for depression varies depending on its severity. Options range from lifestyle changes and self-help strategies to psychotherapy and medication (Beck, 1972). Evidence-based guidelines, such as the NICE guidelines, provide specific recommendations for the treatment of depression (Nice Guidelines, 2022). Cognitive Behavioral Therapy (CBT) is a widely utilized and highly evidence-based psychological intervention for various psychiatric disorders, including depression (Gautam et al., 2020). CBT recognizes the interconnectedness of emotions, thoughts, and behaviors and aims to modify negative patterns by challenging and changing negative thoughts, behaviors, beliefs, and attitudes (American Psychological Association, 2021).

Despite the availability of effective treatment options, a significant proportion of adults diagnosed with depression do not receive treatment (Kessler et al., 2003; Wang et al., 2005). This treatment gap highlights important barriers to accessing care for depression, which will be further discussed in the following sections.

Barriers to treatment

Only around half of individuals with depression receive minimally adequate treatment, such as counseling/psychotherapy or antidepressant therapy (Puyat et al., 2016). Globally, recent data from the World Health Organization (WHO) indicate that only 16.5% of people with major depressive disorder receive treatment each year (Thornicroft et al., 2017). Research has identified various barriers to accessing treatment for depression, including structural, psychological, emotional, and cultural factors (Mitchell & Selmes, 2007; Saraceno et al., 2007).

Structural barriers to treatment access encompass issues such as high costs, limited availability of clinicians, unequal access to healthcare resources, and the heavy workload placed on therapists (Mitchell & Selmes, 2007; Grote et al., 2007; Kohn et al., 2004). Therapists' heavy workloads can lead to limited therapy modalities, longer waitlists, and reduced availability for new patients seeking treatment. In recent years, the waiting times for mental health services have increased, with some individuals waiting several months for treatment (Hähnel et al., 2004; Grünzig et al., 2018). Long waitlists and delays in accessing treatment can be burdensome and may result in individuals giving up on seeking treatment altogether. Mobile health interventions, such as mHealth applications, have the potential to bridge the gap between the desire for treatment and the actual initiation of therapy (Grünzig et al., 2018).

Psychological and emotional barriers to treatment include the stigma associated with depression, skepticism about treatment effectiveness, and a preference for self-help (Kazdin, 2017; Grote et al., 2007; Swartz et al., 2007; Thornicroft et al., 2016). Despite increasing recognition of the biological and physiological nature of mental health conditions, negative views and stigmatization of individuals with mental illness persist (The Lancet, 2016; Thornicroft et al., 2016). Negative self-perceptions and stereotypes surrounding mental illness can also pose psychological burdens.

Cultural barriers play a role as well, with differences in symptom perception and cultural attitudes toward mental illness affecting help-seeking behaviors (Lasalvia et al., 2013; Yousaf et al., 2013). Cultural values and beliefs may influence individuals' willingness to seek help and their acceptance of mental health challenges (Kitano, 1970; C. H. Ng, 1997).

One potential approach to address these barriers is through mHealth interventions, including smartphone applications (Grossman et al., 2020). The use of technology and mobile applications can provide alternative avenues for accessing treatment and support, potentially improving treatment accessibility and overcoming some of the aforementioned barriers. In the following section, the alternative of mHealth interventions will be further discussed.

mHealth applications

Assets and Effectiveness. When considering the barriers to treatment discussed earlier, mHealth applications offer several assets that can help overcome these challenges. In terms of structural barriers, smartphone-based mental health apps can greatly enhance treatment accessibility. Mental health apps have been recognized by prominent public health organizations such as the UK's National Health Service (NHS) and the US National Institute of Mental Health (NIMH) as efficient and adaptable solutions to bridge the treatment gap. The widespread use of smartphones facilitates the exchange of behavioral and health data with healthcare professionals, facilitating personalized treatment plans, real-time progress monitoring, and the ability to make necessary adjustments. (Morris & Aguilera, 2012; Patrick et al., 2008). Additionally, individuals on waiting lists can utilize digital mental health tools, self-help resources, or even engage in remote therapy sessions, providing support during the waiting period and mitigating the negative effects of extended wait times (Grünzig et al., 2018; Patrick et al., 2008).

mHealth options can also help alleviate psychological and emotional barriers. These apps offer convenient and less stigmatizing interventions that may not be available through other means (Morris & Aguilera, 2012; Preziosa et al., 2009). For instance, individuals with mental illness who fear getting diagnosed or receiving treatment due to stigma can find relief with smartphone apps (Kim et al., 2022). The flexibility and anonymity of self-help treatment through these apps can reduce obligations and concerns associated with traditional diagnostic processes or involving third parties, such as health insurance registries (Kim et al., 2022). Research has shown that individuals with significant stigma towards face-to-face treatment may be more willing to explore mental health apps as an alternative option (Kim et al., 2022). Moreover, mHealth apps can cater to individuals from diverse cultures, providing alternative forms of treatment that are less burdened by internalized shame (Bhat et al., 2020).

In addition to addressing treatment barriers, studies have demonstrated the effectiveness of mHealth apps. Various studies investigating the use of mobile phone apps for self-managing depression have consistently shown a correlation between app usage and a reduction in depressive symptoms (Kauer et al., 2012; Morris et al., 2010). A study by Firth et al. (2017)

involving 22 mHealth apps found that using these apps for self-help and symptom reduction resulted in significant symptom reduction compared to a control group. It is noteworthy that cognitive-behavioral therapy (CBT), which is widely recognized as one of the most empirically supported psychological interventions for depression, serves as the foundation for numerous effective mHealth interventions. (Linardon et al., 2019). NICE (National Institute for Health and Care Excellence) guidelines have conditionally recommended digital CBT therapies, acknowledged their effectiveness but also highlighting the need for further research and consideration of specific factors (Nice Guidelines, 2022). It is emphasized that technologies should be used with the support of a mental health professional. Nevertheless, there is much critique on the newly evolving field of mHealth apps. The downsides of mHealth will be introduced in the next section.

Downsides. However, there are downsides to consider in the field of mHealth apps. One significant concern is the lack of scientific research and evidence for most publicly available apps (Lui et al., 2017). Many individuals, without scientific expertise, tend to search for depression apps based on keywords like "depression app" without considering the scientific background or evidence base of the app. The descriptions of these applications often do not reflect the quality of the app, evidence-based functionality, or potential risks (Wasil et al., 2020; Miralles et al., 2020). Marketplaces such as the Google Play Store do not require developers to provide such information, which raises concerns about the lack of an evidence base for many mental health apps (Wasil et al., 2022; Donker et al., 2013).

A study by Stawarz et al. (2018) revealed findings that many highly-rated apps lack a solid evidence base, indicating a discrepancy between user ratings and scientific validity. Ratings on platforms like the Android Google Play Store and Apple App Store are primarily based on subjective user feedback and do not undergo a formal review process or accuracy assessment. Consequently, an app can receive a 5-star rating based on user feedback alone, without any third-party verification of its quality, security, efficacy, or validity (Powell et al., 2016). The lack of consistent oversight or regulation in the mHealth app market allows apps to provide non-evidence-based and inaccurate information without facing any penalties (Martinengo et al., 2019). This absence of reliable oversight contributes to a lack of verification of information quality within these apps.

Currently, there is a notable absence of a formal review process for mHealth apps, which presents a persistent challenge (Powell et al., 2016). This lack of a review process creates difficulties in the adoption and utilization of these apps in healthcare settings or by individuals seeking reliable and evidence-based tools. Moreover, existing medical research on apps is largely deficient in evidence-based approaches and lags behind the widespread utilization of mHealth apps (Powell et al., 2014).

Furthermore, a significant disparity exists between the claims made about app effectiveness in advertisements or app descriptions and the actual evidence-based information available within these apps. This disparity underscores the presence of a "digital research-practice gap," suggesting that the findings from research studies may not necessarily apply to publicly available apps (Wasil et al., 2019; Wasil et al., 2020).

Objective of this review

As was discussed above, depression is a serious illness that affects more people each year. Further, there are many barriers in the treatment process of depression. Therefore, it is no wonder that people try to reach out to self-help options, such as mobile apps. The existence of a gap between the availability and evaluation of depression mobile apps poses a significant problem, as many of these apps are marketed to attract health consumers with unsubstantiated claims of health improvement (Shen et al., 2015), lacking sufficient scientific backing. Although there are many positive factors of mHealth that could help to reduce the barriers to treatment for patients with depression, it is quite difficult to differentiate between the things that are claimed by the developers of the depression apps and the actual evidence-based behind the treatment options of the apps.

Especially a layperson in this field does not know how to scientifically evaluate the available information of these apps or to check whether the treatment elements of the claimed scientific underpinning are really met. Scholars have emphasized the importance of assessing commercial apps for depression in order to facilitate the responsible growth of the expanding market (Shen et al., 2015; Pohl, 2017; Donker et al., 2013). Such evaluations can inform the development of guidelines and standards for app developers, healthcare professionals, and consumers, promoting the responsible use of mental health apps and ensuring their integration into evidence-based care. However, these assessments often concentrate on particular aspects of specific apps, such as ethics considerations and safety, or center on particular therapeutic approaches (Pham et al., 2014; Rumrill et al., 2010), rather than providing a comprehensive evaluation. Hence, the current research paper aims to provide an overview of the state of the art of most popular mHealth applications for people with depression, information the developers of the apps are presenting on evidence-based theoretical background/strategies and if this information holds true. Therefore, the following research questions arise:

What are the key characteristics of top-rated mobile apps that exist for people with depression and do they provide the evidence-based theoretical background/ strategies that are claimed by them?

The research question is divided into the following three questions:

1. What are the top-rated apps for depression and what are their key characteristics/ features?
2. What is their claim on a theoretical background/ strategies?
3. What evidence-based information can be found on the theoretical background/strategies options of the top-rated apps for depression?

Methods

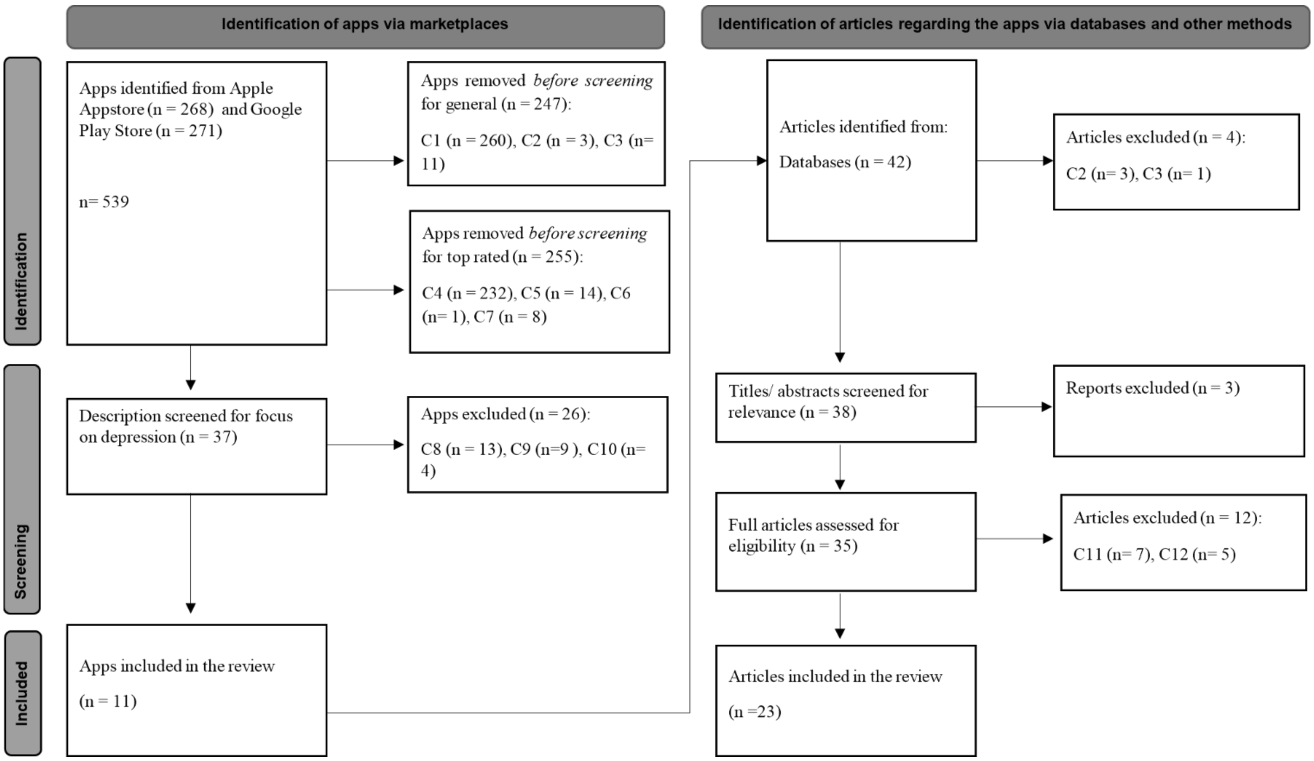
This study adopts a combined approach of an app review and a literature review to comprehensively examine the effectiveness and user experiences of mental health apps for individuals with depression. This research approach aligns with the principles of a scoping review by aiming to provide a broad overview of the available evidence and knowledge on the topic (Pham et al., 2014, Moher et al., 2015).

The app review component involves a systematic search and evaluation of various mental health apps available on popular mobile platforms, such as the Android Google Play Store and Apple App Store. This review captures description, reviews, and ratings of the apps, providing valuable insights into the features, and overall popularity of the apps in addressing depression.

Simultaneously, the literature review component entails a comprehensive search and analysis of scholarly articles, research papers, and relevant sources in the field of mental health and app-based interventions for depression (Munn et al., 2018). This literature review aims to identify and synthesize existing evidence and findings related to the efficacy and theoretical underpinnings of mental health apps for depression.

By combining the app review and literature review, this study adopts a scoping review approach to gather a wide range of information from both user-based perspectives and academic research (Daudt et al., 2013, Levac et al., 2010, Rumrill et al., 2010). This comprehensive approach enables a thorough examination of the current landscape of mental health apps for depression, including their strengths, limitations, and potential implications for clinical practice and future research. The scoping review conducted in this study followed the guidelines outlined in PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) (Tricco et al., 2018). The inclusion of apps and literature in the review was guided by these established guidelines to ensure a systematic and comprehensive approach.

Figure 1



Note. C1 = Criterion 1: Duplicates; C2 = Criterion 2: Not available in English; C3 = Criterion 3: Non available or costs for app; C4 = Criterion 4: Downloaded less than 10.000 times; C5 = Criterion 5: Less than 100 reviews; C6= Criterion 6: Irrelevant categories; C7 = Criterion 7: Average user review scores <4.0 (out of 5.0); C8 = Criterion 8: Does not included “depression” or “depressed”; C9 = Criterion 9: Primary target was not depression; C10: Criterion 10: Single function, with no support or help for depression C10: Criterion 10: Single function, with no support or help for depression; C11: Criterion 11: No focus on one of the selected apps; C12: Criterion 12: No information on theoretical background.

Applications

Search Strategy. The selection of marketplaces for obtaining relevant applications was based on their availability and popularity (Mojica Ruiz et al., 2016). The Apple App Store and Google Play Store were chosen as the primary platforms. The search was conducted between

November and December 2022. To identify the most popular applications, the key terms "depression" and "depressed" were used in App Crawler and Google Play search engines.

The extraction of app data was performed using a script software called Octoparse. Octoparse is a cloud-based web data extraction software that enables users to extract relevant information from various types of websites (Octoparse Data Inc., 2022). In this case, all apps under the search terms "depression" or "depression app" in the app stores were scraped and the unstructured data was saved in Excel (Excel, 2021). Web scraping software allows for the extraction of data from websites or online platforms using the HTTP protocol, similar to web browsers (Sirisuriya, 2015). Web crawlers automate this process and facilitate the extraction of large amounts of data (Zhao, 2017).

During the scraping process, relevant information such as app name, category, price, star review score, number of reviewers, and, in the case of Google Play Store, the number of downloads, were downloaded for each app. This information was recorded in Microsoft Excel, and the total number of reviews from both the Apple App Store and Google Play Store were combined. Additionally, a mean score was calculated for the review scores to facilitate comparison. Initially, the search yielded 539 apps, but after removing duplicates, the final selection consisted of 256 apps.

Selection Criteria. The marketplaces chosen for obtaining relevant applications aimed to include top-rated publicly available apps primarily targeting depression. A selection strategy, outlined in Figure 1, was established to achieve this goal. Initially, general exclusion criteria were set. These criteria included the requirement for apps to be available in English, accessible at the time of selection, and free of charge (as indicated in the app store). Hidden costs such as subscription models or in-app purchases could not be considered due to their display after downloading the app. Additional exclusion criteria were formulated to assess the top-rated apps. Apps were excluded if they had fewer than 10,000 downloads, fewer than 100 reviews, belonged to irrelevant marketplace categories (such as games, social, casual, business, news, or books), or had average user review scores lower than 4.0 (out of 5.0). Applying these exclusion criteria to the initial set of 539 apps resulted in 37 apps for further consideration.

From the resulting 37 apps, a detailed analysis of the app descriptions was conducted, and apps were excluded if they did not primarily focus on the treatment of depression. The exclusion criteria included the absence of the words "depression" or "depressed" in the app store description, the primary target not being depression (e.g., simple diary, sleep schedule monitor, etc.), the app having a single function unrelated to supporting or helping with depression (e.g., wallpaper generator, puzzle, daily quotes, testing tool, etc.), or if the app description explicitly stated that people with depression should not use it. After applying these criteria, a final set of eleven apps (Table 1) was identified for further analysis in the subsequent review.

Data Extraction. Following the extraction of the apps, their information was accessed from both app stores by referring to the app descriptions. The apps were thoroughly reviewed twice to gather descriptive characteristics, which were then summarized in an Excel file (Table 1). The extracted information included the application name, developer, category, number of downloads, target audience, costs, and main function of the app. Additionally, the average rating scores and total reviews were calculated from both the Apple App Store and the Google Play Store.

To address the second research question, a third review of the app descriptions was conducted to determine if the developers claimed their apps to be evidence-based, including explicit scientific foundations and clinical input. Since the average user typically relies solely on the app descriptions provided in the store without seeking further research, the information was extracted solely from there.

To establish a common framework for assessing the evidence base of the apps, the App Rating Inventory (ARI) Checklist (Mackey et al., 2022) was selected. This checklist, which will be further described later, served as a template to examine the claimed evidence base. Specifically, items (a) and (b) in the ARI's "Direct/Indirect Evidence" section were used. For these items, a more general formulation was developed to assess the claimed evidence base. Item (a) involved checking for claimed involvement in formal studies, while item (b) focused on whether the app was claimed to be supported by evidence-based mental health theories in the literature. This allowed for a comparison between the developers' claimed evidence base and the actual evidence base obtained from databases. Additionally, it was examined if the term "evidence base" was mentioned in any of the app descriptions. A summary of the results can be found in Table 2. To gather further information on the actual evidence base of the eleven apps, a second search strategy was developed.

Articles

Search Strategy. The selection of databases for obtaining relevant literature on the eleven applications and their theoretical background was based on their systematic search capabilities and the breadth of literature they cover (Gusenbauer & Haddaway, 2020). For this study, three electronic databases were selected to gather relevant literature: PsycINFO, which is renowned for its comprehensive coverage of the behavioral and social sciences; PubMed, which provides a broad range of biomedical and health-related literature; and Scopus, a comprehensive database encompassing fields such as social sciences, science, technology, and medicine. The inclusion of these databases ensured a diverse and comprehensive collection of literature for the research analysis. The names of the apps were used to search these databases and identify relevant literature. The literature search was conducted in November and December of 2022, resulting in 42 articles.

Selection Criteria. The selected articles needed to be published in English and have full-text availability. The study selection process involved the following steps: first, relevant titles were identified by searching the electronic databases using predetermined search terms. Titles were screened to include either the name of the app or suggest a topic related to mHealth applications that one of the apps could be included in. If the title alone was inconclusive, the abstract of the article was read. Full-text articles were then examined for their relevance. Reports that focused on mHealth applications in general without specific emphasis on one of the selected apps were excluded. Additionally, reports that did not provide any information on the theoretical background of these apps were not included. For example, reports that mentioned the name of an app only as an example without further theoretical background information were excluded. All reasons for excluding articles are illustrated in Figure 1. This process resulted in 23 remaining reports that were used for data extraction.

Data Extraction. After the study selection, the included reports were read, and the relevant data was extracted and summarized in Excel files. This involved capturing the article information, such as the name of the author(s), year, the objective of the article, the study design, and which application was included in the study. These details were compiled and summarized in Table 3.

Quality Extraction of Applications and Articles

To extract the theoretical background/strategies of the selected application and evaluate their quality, a scoring scheme had to be established. The answers to the individual items were summarized in Excel (Table 4).

App Rating Inventory. The App Rating Inventory (ARI) Checklist (Mackey, R., et al., 2022) was chosen as a scoring scheme. The ARI is a tool used to assess mobile applications consisting of 28 items and 3 criteria, utilizing a straightforward binary scoring system. The ARI uses a binary scoring system, where each item is rated as either present (rated as 1) or absent (rated as 0). This approach simplifies the rating process and minimizes subjective biases that can arise when using rating scales with multiple points or levels.

Each app assessed through the App Rating Inventory (ARI) is assigned four scores: evidence, content, customizability, and a total score that represents the sum of the three categories. These scores provide an overall assessment of the app's quality and effectiveness. In this review, the primary emphasis will be placed on assessing the evidential aspects of the apps. The evidence score examines the degree to which the app integrates evidence-based practices and adheres to established guidelines or recommendations (Mackey, et al., 2022). Several factors are taken into consideration, including the inclusion of validated content, references to scientific literature, and adherence to clinical guidelines.

Mackey et al. (2022) conducted a study with the objective of evaluating the App Rating Inventory, a tool developed by the Defence Health Agency. The Agency supports clinical decisions for app selection and the evaluation of medical/ behavioural apps. The ARI was tested on 248 apps with considerably high reliability ($\alpha = .95$). The criterion *Evidence* is subdivided into 3 sections. These sections are *Medical/ Behavioural Focus* with the items: (a) App content focuses on a behavioural or medical concern; (b) Content reflects actual diagnostic nomenclature for the indicated problem; (c) Content is consistent with clinical recommendations/ practice guidelines, *Direct/ Indirect Evidence*: (a) The app itself was involved in at least one formal study (such as a randomized controlled trial); (b) There is reference either in the app store or within the app or can be ascertained by a subject expert, that the app's content is based on the empirical literature;~~,~~ and *Theoretical Model*: (a) The app's content/ interventions are based on an empirically validated treatment model, such as cognitive-behavioural therapy.

Regarding the research question "What evidence-based information can be found on the theoretical background/strategies options of the top-rated apps for depression?", the section Evidence of the ARI was chosen to be the scoring scheme (Mackey, R., et al. 2022). This scheme was used to evaluate the evidence quality from the data extraction of the applications and the result of the literature search. For the gathered data it was checked how many of the beforementioned criterions of the three sections were met for each application. Here, the development study by Mackey, R. et al. (2022) was used as a guideline, as well as common literature to define technical terms. Each section and item will be explained in more detail.

Medical/Behavioural Focus. To assess the medical/behavioral focus of the apps (item a), symptoms of behavioral changes and medical concerns associated with depression were considered, such as withdrawal from social contact, passivity, and loss of daily routines (American Psychiatric Association, 2022; Uher et al., 2014). In cases where apps primarily addressed a concern related to depression without explicitly mentioning the underlying problem, they were evaluated accordingly.

For item (b), the presence of any testing or information related to depression diagnosis criteria was examined. Established instruments like the PHQ-9 depression questionnaire (Kroenke & Spitzer, 2002) were considered, as they assess the DSM-5 criteria for depression.

Clinical practice guidelines (item c) provided evidence-based recommendations for depression assessment and treatment. Directions from the American Psychiatric Association (APA) and the National Institute for Health and Care Excellence (NICE) were consulted, which recommended various psychotherapy interventions and treatment options for depression.

Direct/Indirect Evidence. Formal research (item a) utilizes scientific methods to generate reliable and generalizable knowledge (Holden & Lynch, 2004; Walker, 1997). It follows a systematic approach and involves multiple components to gather essential information and address research objectives (Walker, 1997).

To evaluate empirical literature support (item b), references within the app's description were scrutinized for accuracy, rather than relying solely on information from external websites.

Theoretical Model. To address item (a), the evaluation examined whether the apps were based on empirically validated treatment models of recommended psychotherapy interventions. For example, CBT is rooted in Beck's cognitive model of mental illness (1964), which suggests that emotions and behaviors are influenced by perceptions of events. It was assessed whether the apps went beyond mentioning interventions and included steps that aligned with the underlying treatment model. For instance, if mindfulness was mentioned, the app had to demonstrate its implementation and provide corresponding tools or functions, such as thought diaries or meditation features.

Results

The results section of this paper begins by presenting the characteristics of each application, as outlined in Table 1. Subsequently, the findings from the literature search are introduced and summarized in Table 2. Following the universal characteristics, the results of the claimed evidence found through the app store search are presented in Table 3. Finally, Table 4 provides a comprehensive overview of the actual evidence base of the apps. The findings from both the app store search and literature search are utilized to address the items from the ARI.

Characteristics

The answers to the individual items of the established scoring scheme were summarized in Excel (Table 4).

Applications. A detailed summary of the results of this section can be found in Table 1. Further, the most important findings are presented in the following section.

Table 1

App Characteristics

Application	Developer	Category	Downloads	Mean rating scores	Sum reviews	Target audience	Function of the app
Iona: Mental Health Support	IONA MIND LTD	Health & Fitness	10000	4,6	843	Anxiety, depression	Mood tracker, tools & exercises, personalised plans, journals, meditation, library
Moodfit: Mental Health Fitness	Roble Ridge Software LCC	Medical	50000	4,6	709	Depression and anxiety	Mood journal, goal setting, journal, mindfulness

							& meditation, mood tracker, tests, education
Shmoody: Improve Your Mood	Moodworks Inc.	Health & Fitness	50000	4,5	693	Depression, anxiety	Self-care toolkit, community support
MoodTools - Depression Aid	Inquiry Health LCC	Medical	100000	4,4	3048	Depression	Thought diary, activities, safety plan, information, tests, videos
What's Up? A Mental Health App	Jackson Tempra	Health & Fitness	500000	4,8	3306	Depression, Anger management	Metaphors, diary, habit tracker, catastrophe scale, games, beathing techniques, quotes
Woebot: Your Self-Care Expert	Weobot Labs Inc.	Health & Fitness	500000	4,7	11950	Depression, anxiety	Mood tracker, notifications, practical techniques, chatbot
MindDoc: Your Companion	MindDoc Health GmbH	Medical	1000000	4,5	47090	Depression, anxiety	Mood log, diagnosis, courses and exercises, self help
Remente	Remente	Health & Fitness	1000000	4,1	11694	Anxiety, depression	Tracker, journal, video sessions, goal setting, day planner, courses & activities
Sanvello: Anxiety & Depression	Sanvello Health Inc.	Medical	1000000	4,5	23674	Depression, anxiety, burn out	Mood tracker, scores, activities/courses, coping tools, exercises, community

Wysa: Mental Health support	Touchkin	Health & Fitness	1000000	4,8	125518	Depression, stress, anxiety	AI Chatbot, coping techniques, mindfulness exercises, health reports
Youper: Self-Guided Therapy	Youper, Inc.	Health & Fitness	1000000	4,3	48405	Anxiety, depression, stress	Mood & Symptom tracker, thought journal, CBT sessions

Developer. Most of the applications (9/11, 81%) were developed by companies whose focus was the employment of the app. In most cases (6/11, 56%), the app name was incorporated into the company name (e.g., MindDoc Health GmbH). Only one app (9%) had a sponsor, and Wysa (1/11, 9%) was the only app associated with a separate company. None of the apps received funding or sponsorship from official medical institutes, insurance companies, or other official parties.

Categorization. The eleven apps reviewed in this study were categorized into two categories commonly used to describe apps in marketplaces. The most prevalent category was health & fitness, comprising seven out of eleven apps (64%), followed by medical apps with four out of eleven (36%).

Downloads. Among the reviewed apps, only one app (9%) had approximately 10,000 downloads, while another app (9%) had 100,000 downloads. Two apps (18%) had 50,000 downloads, and two apps (18%) were downloaded 500,000 times. Five apps (45%) were downloaded over 1,000,000 times. The average mean rating score for the eleven apps was 4.5%. The number of reviews for the apps varied from 693 reviews (Shmoody) to 125,518 reviews (Wysa).

Costs. All the reviewed apps were initially free to download. However, users may incur costs indirectly through forced consumption of in-app advertisements or explicit charges for more advanced features. Due to limitations in the research methodology, the presence of in-app purchases could not be assessed as it required downloading and using the applications. Hence, this aspect could not be included in the evaluation process.

Target Audience. All the included apps claimed to target users with depression. Additionally, the majority of the apps also targeted anxiety (8/11, 72%) and stress (5/11, 45%), as well as other mental health issues (7/11, 64%). None of the apps specifically targeted users with a particular level of severity (i.e., mild, moderate, or severe depression)

Function of the App. The functions of mood or thought tracking were included in 63% (7/11) of the apps. Another 63% (7/11) of the apps featured goal setting functions, planners, or personalized plans. Five apps (45%) offered a journal or diary. Exercises and activities were included in 45% (5/11) of the apps. Approximately 45% (5/11) of the apps provided self-help options, which were not further specified by the developers but labeled as "practical techniques." Meditation and mindfulness were featured in 36% (4/11) of the apps. Around 27% (3/11) of the apps offered a library with information on the topic, while 18% (2/11) included additional educational videos. Two apps (18%) incorporated an AI chatbot, and three apps (27%) provided a test function. Other functions found in the apps were community support (2/11, 18%), metaphors and quotes (2/11, 18%), and games (1/11, 9%).

Articles. A comprehensive literature search yielded a total of 23 articles relevant to the topic. The retrieved articles encompassed various research methodologies, with 35% being systematic reviews (8/23), 17% consisting of randomized controlled trials (4/23), and 9% representing cross-sectional studies (2/23). Additionally, qualitative research methods were employed, resulting in two thematic analyses (9%). Each of these scholarly articles focused on one or more of the applications identified in the marketplace search. Notably, Wysa emerged as the most extensively studied app, with eight articles (35%) dedicated to its evaluation. Woebot received significant attention as well, being the subject of seven articles (30%). Other noteworthy apps included MoodTools, which was discussed in 5 articles (22%), and Youper, which appeared in four articles (17%). For Sanvello, What's Up?, and MindDoc 2 (9%) articles were found and Moodfit was represented with one (4%) article. However, no literature was available for Iona, Shmoody, and Remente, indicating a paucity of academic research on these applications.

Table 2

Article Characteristics

Author, Year	Objective	Design	Applications included
Wasil, A. R., et al., 2021	Review highly popular self-help apps for depression and anxiety	Systematic review	Wysa
Wasil, A. R., et al., 2019	Assess the extent to which evidence-based treatment elements are present and their frequency within popular smartphone apps for depression and anxiety	Systematic review	Wysa, MoodTools, Youper
Huguet, A., et al., 2016	Identify and evaluate CBT or BA self-help apps on their usefulness, usability, and integration and infrastructure	Systematic review	MoodTools
Sinha, C., Cheng, A. L., & Kadaba, M., 2022	Evaluate user retention and engagement with an artificial intelligence–led digital mental health app (Wysa).	Thematic Analysis	Wysa

Moberg, C., Niles, A., & Beermann, D., 2019	This study sought to validate the effectiveness of Pacifica, a popular commercially available app for the self-management of mild-to-moderate stress, anxiety, and depression.	Randomized Waitlist Controlled Trial	Sanvello
Su, L., & Anderson, P. L., 2022	Examine the naturalistic user behavior of MoodTools.	Observational study in a global sample (N= 158930)	MoodTools
Stawarz, K., et al.	Analyze functionality and user opinions of mobile apps purporting to support cognitive behavioural therapy for depression and to explore key factors that have an impact on user experience and support engagement.	Systematic Review	MoodTools
Rebedew, D., 2018	Guide patients to the most helpful apps, this article presents five that stand apart from the rest when reviewed using FPM's “SPPACES” criteria.	Medical App Review	MoodTools, What’s Up?
Darcy, A., et al., 2021	Investigate whether users of a CBT–based conversational agent would report therapeutic bond levels that are like those in literature about other CBT modalities.	Cross-sectional study (N= 36070)	Woebot
Darcy, A., et. al., 2022	Introducing AGCBT as a new model of service delivery, whilst describing Woebot.	Expert Review of Medical Devices	Weobot
Carlo, A. D., Ghomi, R. H., Renn, B. N., Strong, M. A., & Areán, P. A. (2020)	Examined download and utilization data of behavioural health apps with a focus on stickiness	Cross-sectional study	Youper

Prochaska, J. J., et al., (2021)	Compared usage of Woebot for 8 weeks to a waitlist control	Randomized controlled trial (N=180)	Woebot
Monnier, D., 2020	Examine the promise of what is “presented as the future of psychotherapy”	Systematic Review	Weobot
Fitzpatrick, K. K., Darcy, A., & Vierhile, M., 2017	Determine the feasibility, acceptability, and preliminary efficacy of a fully automated conversational agent to deliver a self-help program for college students with anxiety or depression	Randomized controlled trail (N= 70)	Weobot
Beatty, C., et al., 2022	Examine whether users perceive a therapeutic alliance with an AI conversational agent (Wysa) and observe changes in the therapeutic alliance over a brief period.	Mixed-methods study (N= 1205)	Wysa
Malik, T., Ambrose, A. J., & Sinha, C., 2022	Analyze feedback content to understand users’ experiences with engaging with a digital mental health app and capture the types of mental health app users.	Qualitative Thematic Analysis	Wysa
Inkster, B., Sarda, S., & Subramanian, V., 2018	Present a preliminary real-world data evaluation of the effectiveness and engagement levels of an AI-enabled, empathetic, text-based conversational mobile mental well-being app, Wysa, on users with self-reported symptoms of depression.	Mixed Methods study	Wysa
Nicol, G., Wang, R., Graham, S., Dodd, S., & Garbutt, J., 2022	Evaluated the feasibility of delivering the app-based intervention to adolescents with	Feasibility and Acceptability Study	Woebot

	moderate depressive symptoms who were treated in a practice-based research network (PBRN) of academically affiliated primary care clinics.		
Li, L. S. E., Wong, L. L., & Yap, K. Y. L., 2021	Evaluate the quality of stress, anxiety and depression apps recommended for COVID-19.	Systematic Review	Sanvello, Moodfit, What's Up? Wysa, Youper, Woebot
Voderholzer, U., et al., 2021	Examine changes in depressive symptoms and life satisfaction during outpatient care (MindDoc) for patients with depressive disorders, the quality of the established working alliance, and the influence of working alliance and the patients' technology commitment.	Study (N= 59)	MindDoc
Kerber, A., et al., 2021	Examine improvement of app-based unguided self-management mental health literacy, patient empowerment, and access to care for people with mental health impairments	Randomised controlled trial (N= 1100)	MindDoc
Mehta, A., et al., 2021	Examine the acceptability and effectiveness of Youper	Longitudinal observational study (N= 4517)	Youper
Meheli, S., Sinha, C., & Kadaba, M. (2022)	evaluate the perceived needs, engagement, and effectiveness of the mental health app Wysa regarding mental health outcomes among real-world users	Mixed Methods Retrospective Observational Study (N=2194)	Wysa

Claimed Evidence Base in the Appstore description

The claimed evidence was assessed based on the criteria developed earlier and examined in the descriptions provided on app stores (Table 3). However, the available information regarding involvement in studies was limited. Only Woebot and Youper (18%) mentioned their participation in scientific studies. On the other hand, ten out of eleven (91%) applications included statements in their descriptions suggesting support from literature or expert involvement. Moodfit was the only app where no information regarding evidence or studies was found. The extracted quotes from the app stores are presented in Table 3, which were consistent across both the Apple Store and the Google Play Store. Notably, three out of eleven (27%) apps, specifically MindDoc, Woebot, and Wysa, mentioned the keyword "evidence-base" in their descriptions.

Table 3

Claimed Evidence Base by the Developers

Application	Claimed involvement in scientific study	Claimed back up by literature/ Expert involvement	Mentioned keyword “evidence based”
Iona	N/A	“Scientifically backed up tools from CBT”	N/A
MoodTools	N/A	“Collaboration with multiple mental health professionals”, “Contains several different research-supported tools	N/A
Moodfit	N/A	N/A	N/A
MindDoc	N/A	“Developed by clinical psychologists, leading researchers about emotional well-being", “Content aligned with international treatment guidelines for mental disorders”	“...providing both transdiagnostic and disorder-specific evidence-based courses and exercises.”
Remente	N/A	“... written by psychologists, business managers, life coaches and world champions in several fields”	N/A

Sanvello	N/A	"designed by experts", "customized tools, rooted in cognitive behavioural therapy", "concepts of CBT"	N/A
Shmoody	N/A	"clinical research that backs it up"	N/A
What's Up?	N/A	"the best CBT (Cognitive Behavioural Therapy) and ACT (Acceptance Commitment Therapy) methods"	N/A
Woebot	"Studies show that it works. In a clinical trial involving 400 participants, Woebot users showed a 32% reduction in depression and a 38% reduction in anxiety after just four weeks. Check out more of Woebot's clinical results at https://woebothealth.com/clinical-results/."	"using proven therapeutic frameworks like Cognitive Behavioural Therapy (CBT)" "guides you through practical techniques based on tried and tested approaches such as Cognitive Behavioural Therapy (CBT)" "Crafted by Experts: Born out of research led by clinical research psychologist Dr. Alison Darcy, Woebot has demonstrated efficacy in published randomized controlled trials."	"Woebot was built on a foundation of clinical evidence base"
Wysa	N/A	"Research-backed, widely used techniques of cognitive behavioural therapy (CBT)",	"100+ evidence-based stories from clinical team"
Youper	"A study from Stanford University showed significant improvement in symptoms of depression and anxiety after using the Youper app."	"Youper uses Cognitive Behavioural Therapy (CBT) techniques, the scientifically proven way to improve your mood.", "Youper was created by therapists to make CBT accessible for everyone"	

Actual Evidence Base

The actual evidence was assessed by analyzing the information obtained from the app stores and literature search, and then evaluated using the ARI. The average score for the evidence-based sections of the ARI across the eleven apps was M = 3.9. Youper had the highest number of evidence-based items (5, 45%), while Remente and Shmoody had the lowest number (1, 9%). Table 4 provides a comprehensive overview of the results for each ARI item, and additional information on Direct/Indirect Evidence can be found partially in Table 2 under *Design* and *Applications Included.*

Medical/Behavioural Focus. The findings for item (a) revealed that each app addressed behavioral or medical concerns. Table 1 section *Target audience* already provides an overview of the applications and their respective concerns, found in the app description. The literature search confirmed this observation and identified additional concerns.

All apps acknowledged the concern of feeling depressed or having a negative mood. Additionally, feeling anxious was a prevalent concern in 91% of the apps (10/11). Six apps (55%) addressed stress as a concern, while poor sleep or insomnia was mentioned in 45% of the apps (5/11).

Regarding item (b), only Iona, MoodTools, and Moodfit (27%) incorporated diagnostic nomenclature for the indicated problem, specifically the depression test questionnaire PHQ-9.

In terms of item (c), eight apps (73%) incorporated clinical recommendations or practice guidelines, often involving cognitive-behavioral therapy (CBT). Among these eight apps, four also included additional psychological interventions. Moodfit included mindfulness (9%), Wysa incorporated Dialectical Behavior Therapy (DBT) (9%), and Whats's Up? included Acceptance and Commitment Therapy (ACT) (9%).

Direct/Indirect Evidence. The results for item (a) *The app itself was involved in at least one formal study* and item (b) *Reference either in the app store or within the app or can be ascertained by a subject expert, that the app's content is based on empirical literature* are going to be presented for each application that information was available for in the following. If no context is given to answer the item, the information was not available. Overall, six out of eleven (55%) app were involved in at least one formal study. Weobot was involved the most, with four studies. For item (b), only 18% (Weobot, Youper) had a reference in the app store.

MoodTools Though included in an observational study with a large sample size (N= 158,930), no significant findings were reported (Su, L., & Anderson, P. L., 2022).

MindDoc In a randomized controlled trial with over 1000 participants, the app demonstrated significant improvements in attitudes towards mental health, self-management behaviors, healthcare utilization, psychopathology, and quality of life for patients with various mental disorders (Kerber, A., Beintner, I., Burchert, S., & Knaevelsrud, C., 2021).

Sanvello. In a randomized study of 500 adults with anxiety and depression, the tools of Sanvello were shown to decrease symptoms, with sustained effects even after discontinuing app usage (Moberg, C., Niles, A., & Beermann, D., 2019).

Woebot. Several formal studies found Woebot to be effective and feasible in reducing depression and anxiety symptoms, as well as in supporting individuals with substance use disorders (Fitzpatrick et al., 2017; Ramachandran et al., 2020; Prochaska et al., 2021; Vogel et al., 2021).

For item (b), Weobot offered a reference for their evidence-based and a link to their webpage, which leads to the list of studies that are described above.

Wysa. Included in three studies, Wysa demonstrated clinically meaningful improvements in depression, higher engagement levels, and comparable outcomes to traditional CBT approaches (Leo, A. J., et al., 2022; Inkster, B., Sarda, S., & Subramanian, V., 2018; Beatty, C., Malik, T., Meheli, S., & Sinha, C., 2022).

Youper. Two studies showed that Youper users experienced reduced depression symptoms within a short period of app usage and highlighted the inclusion of cognitive behavioral therapy skills in the app's content (A., Niles, A. N., Vargas, J. H., Marafon, T., Couto, D. D., & Gross, J. J., 2021; Carlo, A. D., Ghomi, R. H., Renn, B. N., Strong, M. A., & Areán, P. A., 2020).

For item (b), the app store description of Youper mentions a foundation of clinical evidence and a statement that clinical studies prove that claim. Afterwards, a reference is given for that claim and a link to the website of Youper. The linked studies are described above.

Theoretical Model. The following results are going to present if the treatment model of the psychological intervention from the section *medical/behavioural focus* item (c) is supported by the tools of the app. First, it is shown if it is mentioned how the psychological intervention was implemented or supported by a treatment model. Next, the tools of the application are presented.

Iona. The app is built upon the principles of Cognitive Behavioral Therapy (CBT). It offers various tools and features that align with CBT, including mood tracking, identification of thinking patterns and behaviors, exercises to challenge patterns, personalized plans, gratitude journaling, and meditation.

MoodTools. The app claims to be based on CBT but does not provide specific details on the implementation. It includes a thought diary for analyzing thoughts and identifying negative thinking patterns. However, more information is needed to determine if these patterns are challenged. The app also offers activities, a suicide safety plan, self-help guides, and educational videos.

Moodfit. This app states to be based on both CBT and mindfulness. It supports users in identifying distorted thinking through thought records and offers tools like a mood journal, daily goals (including gratitude and mindfulness practices), and breathwork.

Sanvello. Although Sanvello mentions the inclusion of CBT tools and changing thoughts and behaviors, it does not specify how this is implemented. The app offers therapy, coaching, coping techniques, meditations, and mood tracking, but lacks explicit support for a validated CBT treatment model.

What's Up?. This app provides methods to overcome common negative thinking patterns and offers tools such as a comprehensive diary, metaphors, a catastrophe scale, a grounding game, positive quotes, and breathing techniques. It mentions the implementation of CBT and its tools support the cognitive model of depression.

Woebot. Woebot is grounded in CBT, but the exact implementation is unclear. It features lessons, interactive exercises, videos, mood tracking, and skills covering various topics. However, more information is needed to assess how CBT is incorporated into the app.

Wysa. The app does not describe the implementation of CBT techniques clearly. It mentions an AI chatbot that teaches CBT techniques, as well as meditation, yoga, and guided journaling activities. Insufficient information is available to evaluate the criteria.

Youper. Youper implements CBT by analyzing thoughts, challenging negative thoughts, and promoting alternative ways of thinking. It offers tools such as problem-solving, goal setting, thought journaling, mood tracking, and symptom monitoring for anxiety, depression, and other mental health symptoms. The app informs about CBT implementation and provides supporting tools.

Overall, the apps varied in their level of clarity and explicit support for the respective theoretical models.

Table 4

ARI Evidence Base for the applications

Application	Medical/ Behavioural focus item (a)	Medical/ Behavioural focus item (b)	Medical/ Behavioural focus item (c)	Score	Direct/ Indirect Evidence (a)	Direct/ Indirect Evidence (b)	Score	Theoretical Model (a)	Sum Score
Iona	Low mood, poor sleep, feeling disconnected Stress, Anxiety, and depression	PHQ-9	CBT	3/3	N/A	N/A	0/2	Yes	4/6
MoodTools	Sad, anxious, or depressed, negative moods	PHQ-9	CBT	3/3	1	N/A	1/2	N/A	4/6
Moodfit	Stress, struggling, procrastination, rumination	PHQ-9	CBT, Mindfulness	3/3	N/A	N/A	0/2	Yes	4/6
MindDoc	depression, anxiety, insomnia, and eating disorders	N/A	N/A	1/3	1	N/A	1/2	N/A	3/6

Remente	Sleep problems, stress anxiety, depression	N/A	N/A	1/3	N/A	N/A	0/2	N/A	1/6
Sanvello	anxious, lonely, burned out	N/A	CBT	2/3	1	N/A	1/2	Partially met	3/6
Shmoody	Depression, anxiety, feeling down	N/A	N/A	1/3	N/A	N/A	0/2	N/A	1/6
What's Up?	depression, Anxiety, Anger, Stress and more	N/A	CBT, ACT	2/3	N/A	N/A	0/2	Yes	3/6
Woebot	Sad, anxious, stress	N/A	CBT	2/3	4	App Store	2/2	N/A	4/6
Wysa	depression, stress, anxiety, sleep and a whole range of other mental health and wellness needs	N/A	CBT, DBT	2/3	3	N/A	2/2	N/A	3/6
Youper	Anxiety, Depression, Insomnia, Substance addiction	N/A	CBT	2/3	2	App Store	2/2	Yes	5/6

Discussion

An app review on the key characteristics and evidence base of top-rated mobile apps that exist for people with depression with an additional literature review on the theoretical background/strategies of these apps was conducted. Here, the focus laid on what the producers claim of the app and what actual evidence-based background could be found in app stores and literature. For the second part, the ARI was used to evaluate the emerged eleven top-rated mobile apps based on their actual evidence.

Findings, Indications & Implications

Characteristics. In response to the first sub-question of the research inquiry, the app review yielded a total of eleven applications. The key findings on their characteristics/features will the presented in the following.

App review. For the app review, it was the goal to establish the top-rated apps for depression. After an app search and review on the ground of developer, category, number of downloads, target audience, costs, main function of the app, reviews, and mean ratings scores, eleven top-rated apps were established. These apps still greatly varied in number of downloads between 10.000 and 1.000.000. The distribution of app downloads, varying from 10.000 to 1.000.000, among the top-rated apps in app stores can be attributed to various factors. These include variations in popularity, marketing strategies, app features, user preferences, positive reviews, word-of-mouth recommendations, and the level of competition in the market. Other research suggests that there are severe variations in app usage (Wasil, Gillespie, Shingleton, et al., 2020). Furter, the study by Wasil et al. (2020) investigated the usage patterns of different mental health apps and found that two specific meditation apps, Headspace and Calm, were disproportionately popular among users seeking apps for depression. In fact, these two apps accounted for more than 50% of the user base for depression apps. This information indicates that a small number of apps, in this case, Headspace and Calm, dominate the market or user preferences within a specific mental health category. It could imply that these apps have gained a significant reputation or recognition for addressing depression-related concerns, leading to a larger user base compared to other similar apps (Shingleton, et al., 2020). This factor should be taken into consideration when conducting app reviews, particularly when evaluating top-rated apps, as the usage pattern can provide valuable insights.

When looking at the findings regarding the app developers, it is evident that none of the apps established partnerships with official medical institutes, insurance companies, universities, or other authoritative parties. This could indicate a lack of external validation, as the app's claims and effectiveness may rely solely on research funded by the developer (Azhar et al., 2015). Secondly, it highlights limited access to resources and expertise that could potentially enhance the app's quality and credibility (Luborsky et al., 1999). Moreover, it raises concerns about the app prioritizing commercial interests over providing unbiased mental health support (Grundy et al., 2016)

In the reviewed apps, a wide range of medical concerns were addressed; however, none of them specifically targeted users based on the severity of their condition, such as mild, moderate, or severe depression. This should be viewed critically because the needs and treatment approaches vary significantly depending on the severity of the condition. What works for individuals with mild depression may not be sufficient or appropriate for those with moderate or severe depression. The study conducted by Firth et al. (2017) further corroborates this finding by showing that substantial improvements were only observed in individuals with self-reported mild-to-moderate depression who used smartphone apps, highlighting the importance of considering the specific target population type these apps. For example, someone with severe depression might need more intensive interventions, such as therapy or medication management, while someone with mild depression might benefit from self-help techniques or guided interventions. By not addressing severity levels, users may not receive the appropriate level of support that aligns with their needs.

Lastly, the abundance of apps available when searching for depression apps in the app store should be mentioned. Upon scrolling past the initial recommended apps, it became evident that there was a high quantity of apps solely focused on providing one single task, such as daily quotes or "depression wallpapers" for the phone. This reflects the fact more than 350,000 digital health apps are currently available to consumers (Pohl, 2023). Although, while a funny quote each morning might indeed lift some people's mood, these apps lacked any indication of having a theoretical background or evidence-based approach. As an observer, I noticed distinct differences among the various apps, but it remains questionable whether non-professionals would be aware of these distinctions. This observation aligns with previous studies that have examined a wide range of apps on commercial app stores, revealing significant disparities in quality between the top-ranked apps and the majority of available apps (Wasil et al., 2020). The concern here is that non-professionals may struggle to differentiate between the apps, they may not be able to decide which apps are the right ones for their issue and might not scroll past the first few recommended ones.

Literature search. The findings from the literature search revealed that among the evaluated apps, Wysa stood out significantly as the most extensively studied app. The fact that Wysa has the most studies conducted on it indicates that Wysa has received extensive attention from researchers, demonstrating its perceived scientific importance by either the company itself or others (Ahmad et al., 2018). The multiple studies conducted on Wysa enhances the app's credibility and fosters confidence among users and healthcare providers (Firth et al., 2017, Schnall et al., 2016). However, it's important to note that further examination of the specific research findings is necessary to determine the reliability and generalizability of the results.

Furthermore, the absence of scientific studies on five apps, suggests that they have no research to validate their claims or evaluate their outcomes, these apps have limited or no evidence to demonstrate their theoretical background in addressing mental health concerns (Schnall et al., 2016). This findings may initially appear not much, but when considering the broader scheme, they present a positive outlook. Notably, the study conducted by Lau et al. (2020) revealed that only 2% of the 1,000 publicly available mHealth apps had undergone scientific testing. However, the findings of this review differ significantly, as 55% of the reviewed apps were found to have been part of a scientific study. This contrast could indicate among other things, which will be introduced further into the discussion, an improvement in the inclusion of scientific research within the development and evaluation of mental health apps.

Evidence Base. The subsequent section will primarily focus on the key findings related to the second and third sub- research questions. Specifically, the discussion will revolve around the claims made by the applications regarding their evidence base and the actual supporting information found to substantiate these claims.

When comparing the findings above to the claimed evidence, almost all applications claimed of being backed up by literature/expert involvement with phrases such as “Scientifically backed up tools” or “clinical research that backs it up”. Yet, when comparing the claims to the actual evidence base, only six apps were involved in at least one formal study. This aligns with other studies that have found that a significant number of apps have not been rigorously evaluated or tested in controlled research settings (Torous et al., 2017). Based on this finding, it can be concluded that there is a disparity between the claims made by the app developers regarding theoretical background and strategies, and the actual evidence base supporting those claims (Nicholas et al., 2019) A lack of transparency and potentially misleading marketing practices by app developers who make bold claims about their apps being scientifically backed or supported by clinical research, despite limited or no evidence to substantiate those claims. This can mislead users and hinder their ability to make informed decisions about the efficacy and reliability of the apps (Torous et al., 2017). Two apps, namely, Remente and Schmoody, which both were not included in any study, also received the lowest scores on the overall evidence-based items on the ARI. It implies that apps without research backing may lack sufficient evidence to support their effectiveness and reliability (Buijink et al., 2013).

Contrary to the paragraph above, it is worth mentioning that both Woebot and Youper explicitly stated in their app descriptions that their content was based on empirical literature, and these claims were substantiated. Further, Woebot claimed involvement in only one study in the description, while further investigation revealed its participation in three scientific studies. On the other hand, Wysa did not make any explicit claims regarding involvement in scientific studies in the app description, yet the literature search revealed its participation in three studies. This raises questions about their emphasis on research and missed opportunities to highlight their evidence base. It is questionable why this information was not highlighted. However, the existence of these studies demonstrates that the app has undergone some level of scientific investigation, which enhances its credibility and potential effectiveness (Buijink et al., 2013).

Furthermore, the average score of 3.9 out of 6 on the App Review Index (ARI) for evidence base indicates an above-average performance. Contrary to the initial expectations, the applications demonstrated a stronger evidence base than anticipated, considering previous research highlighting the lack of theoretical background in publicly available mental health apps

(Lau et al., 2020; Lui et al., 2017; Donker et al., 2013). It is important to note that the selection of apps in this review and the specific criteria used may have influenced the outcome, leading to different findings compared to previous studies that employed varied methodologies and assessed a broader range of apps (Lui et al., 2017; Donker et al., 2013). The evaluation of mental health apps is a multifaceted process that encompasses user experience, app functionality, and evidence base (Le et al., 2021). While the present research adds new insights, it underscores the ongoing need to enhance the evidence base and theoretical foundations of mental health apps. Continued efforts are essential to advance the quality and evidence base of these applications.

When looking at the results of the ARI, Youper achieved the highest score for evidence base overall. This is further supported by its involvement in two studies and the incorporation of CBT and related tools within its app. Therefore, combination of a measurably strong evidence base, research involvement, and utilization of established therapeutic approaches demonstrates Youper's commitment to providing an evidence-based mental health solution.

Furthermore, upon closer examination of the results, an important aspect warrants discussion. Despite the increased involvement in studies and the above-average performance indicated by the ARI results, it is necessary to critically evaluate the findings regarding the inclusion of a theoretical model. While 73% of the apps incorporated Cognitive Behavioral Therapy (CBT) in their design, it is noteworthy that only four apps provided comprehensive information on the implementation of CBT. The limited provision of detailed information on CBT implementation, despite claiming its incorporation, raises concerns about the app's theoretical foundation, adherence to evidence-based practices, and reliability of CBT components (American Psychological Association, 2021; Gautam et al., 2020). This finding aligns with previous research, which has also highlighted the limited integration of CBT principles in publicly available mHealth applications (Donker et al., 2013; Shen et al., 2015). Thus, the findings of this review reinforce the existing deficiency in integrating CBT principles in these apps and show a greater need for transparent and integrated CBT principles in mental health apps is necessary for their effectiveness and credibility.

Finally, it should be pointed out that four apps, namely Moodfit, What's Up?, Youper, and Iona, demonstrate a notable alignment with a theoretical model, specifically Cognitive Behavioral Therapy (CBT). These apps not only incorporate CBT principles but also offer a range of tools that correspond to those principles. This indicates a systematic and evidence-based approach to mental health support, which is a positive finding despite their performance on other variables (Gautam et al., 2020). Among them, Youper stands out with the highest score on the ARI, research involvement, and the utilization of CBT principles. This combination indicates Youper's dedication to providing an evidence-based mental health solution, highlighting the importance of a strong evidence base, research involvement, and the utilization of established therapeutic approaches in mental health apps.

Limitations and Recommendations

The present review provides valuable insights into the landscape of top-rated depression apps and their evidence base. However, it is important to acknowledge certain limitations and consider recommendations for future research and practice. This section outlines the limitations encountered during the study and offers recommendations to address these limitations and further enhance the development and evaluation of mental health apps.

The first limitation to address is related to the exclusion criteria applied in the current study. Specifically, the criterion requiring the inclusion of "depression" in the app store description resulted in the exclusion of some of the most widely downloaded mental health apps, such as Headspace and Calm. These two apps account for a significant portion of the user base for depression apps, comprising over 50% of the total users. Further support for this observation can be found in a study conducted by Wasil et al. (2020), which indicated that the top three apps for depression accounted for approximately 90% of downloads. In the study, the three apps were Headspace, Youper and Wysa , which two are also topic of the current review. Further, the apps were in fact the two apps with the highest downloads (>1.000000). Consequently, this exclusion may have contributed to the observed wide distribution of downloads and reviews mentioned earlier in the analysis. While the inclusion of "depression" in the app store description provided a specific focus on depression-related apps, it also resulted in the exclusion of popular mental health apps with a broader user base.

To enhance the representativeness of the study, future research could consider a more inclusive approach that encompasses a wider range of mental health apps, including those without explicit mention of depression. Additionally, future studies could explore the effectiveness of the excluded apps, such as Headspace and Calm, in the context of depression management. This would provide valuable insights into the potential benefits of these widely used apps for individuals experiencing depressive symptoms. Lastly, it could be beneficial for future research to investigate the specific reasons behind the high spread distribution of downloads and reviews observed in the included apps.

The use of the App Review Index (ARI) as the scoring tool for assessing the evidence base of the reviewed apps introduces a potential limitation to this study. It is important to acknowledge that the ARI is not widely used and there is limited research available on its validity and reliability (Mackey, R., et al., 2022). The decision to employ the ARI was a delicate one, considering the scarcity of established instruments for scoring the evidence base of mental health apps. Choosing the ARI posed a risk, but it was driven by the lack of alternative tools specifically designed for evaluating the evidence base of apps. With limited existing research in this area, the selection of an appropriate scoring tool becomes challenging. The ARI, although less extensively validated, offered a structured framework to assess key aspects of the evidence base in a systematic manner, without having to download the app and evaluate the user interface. Yet, it is essential to acknowledge its limitations and the need for further validation and refinement of scoring instruments tailored to evaluate the evidence base of mental health apps accurately with few research tools available.

Further, the utilization of the ARI also impacted the extent of the review, necessitating detailed explanations of each individual item. As a future recommendation, a comprehensive review could be dedicated solely to the refinement and expansion of the ARI or other scoring methods. Recently, the APA has started to develop a standard for evaluating mental health apps for consumers and clinicians (American Psychological Association, 2021). Such an endeavour would aid researchers with limited access to research tools and even non-professionals in conducting straightforward evaluations of the evidence base of apps. The implications extend to the wider research community and non-professionals interested in evaluating app quality. The development and validation of user-friendly scoring instruments can empower researchers with limited research tools and enable non-professionals to assess the evidence base of apps more easily. This inclusivity and accessibility have the potential to foster greater transparency, accountability, and user empowerment in the digital mental health landscape.

Another limitation that emerged considering the limited tools in the research methodology employed in this review was the inability to assess in-app purchases. This limitation arose because evaluating in-app purchases would have required downloading and actively using the applications, which was not within the scope of this study. This limitation could give rise to potential issues as in-app purchases hold significant importance in numerous mobile applications, exerting considerable influence on the user experience, app functionality, and overall value. In-app purchases encompass a wide range, spanning from ad removal and unlocking additional features to accessing premium content or virtual goods. Gaining an understanding of the presence and characteristics of these purchases is vital for conducting a comprehensive evaluation of the financial implications for users. Neglecting to assess in-app purchases may result in overlooking crucial insights into the app's revenue model, user experience, and potential limitations. Based on the limitation, it is recommended that future research in this area addresses the assessment of in-app purchases in mobile applications.

Future research focused on the evidence base of mental health apps could benefit from conducting more detailed evaluations across various areas. By delving deeper into specific aspects, researchers can gain a more comprehensive understanding of the evidence base and its implications for app effectiveness and user outcomes. In light of this consideration, it is important to acknowledge another limitation concerning the available research on mental health apps. While six applications were included in studies, the emphasis was not placed on evaluating the quality of the literature or conducting further analysis of the results. This highlights a potential limitation of your study in terms of the comprehensiveness of the evidence base assessment. While the inclusion of apps in studies provides some level of credibility and indicates a certain level of scientific investigation, the lack of critical evaluation and detailed analysis of the literature limits the depth of understanding regarding the effectiveness of these apps. This limitation should be acknowledged when interpreting and generalizing the findings of your study. Secondly, it underscores the need for future research to not only consider app inclusion in studies but also to conduct rigorous evaluations of the quality of the literature and thoroughly analyze the results. By placing more emphasis on critically evaluating the evidence base and examining the specific findings, researchers can gain deeper insights into the effectiveness and reliability of the apps.

Lastly, the scope of this review predominantly focused on Cognitive-Behavioral Therapy (CBT) as the primary theoretical model. This emphasis was driven by its frequent mention in previous literature on depression mHealth apps (Nice Guidelines, 2022; Mitchell & Selmes, 2007; Linardon et al., 2019; Gautam et al., 2020). CBT also received significant attention during the app and literature review stages. It is important to bear this in mind when evaluating the results.

However, it is important to acknowledge that there are other evidence-based interventions for depression worth considering. For example, mindfulness-based therapies such as Mindfulness-Based Stress Reduction (MBSR) and Mindfulness-Based Cognitive Therapy (MBCT) that integrate mindfulness meditation and practices that promote non-judgmental awareness of the present moment (Khoury et al., 2013). Given that this review revealed the presence of mindfulness tools within the reviewed apps, further exploration of these options in the context of depression apps would be worthwhile.

Conclusion

The current scoping review has provided valuable insights into the claimed and actual evidence base of top-rated mHealth applications for depression. While this review successfully addresses the research question by exploring the key characteristics of the top-rated depression apps within the scope of the review, providing a definitive answer to the second part is challenging. However, a more detailed assessment, particularly regarding the implementation of

Cognitive Behavioral Therapy (CBT), reveals a gap between developer claims and the actual evidence base. Closing these gaps will contribute to the advancement and reliability of mental health apps, ultimately benefiting users seeking evidence-based support.

Among the reviewed apps, only Moodfit, What's Up?, Youper, and Iona provide an evidence-based theoretical background/strategies as claimed. Despite Wysa being the one app with the most research attention. Notably, Youper emerges as the app that fully aligns with its claims, exhibiting a medical/behavioral focus, direct and indirect evidence, adherence to a theoretical model, and the highest score on the ARI. Overall, the review exposes a discrepancy between developer claims or advertising and the actual evidence base. However, it is noteworthy that the apps include more evidence-based background than initially anticipated. While this research provides new insights into the evidence base of apps, it underscores the ongoing need to enhance the evidence base and theoretical foundations of mental health apps. Continuous efforts are essential for advancing the quality and evidence base of these applications.

Moreover, the review highlights the need for detailed and thorough research, including studies and in-app evaluations, to further explore the evidence base of these apps. Additionally, the development of valid and accessible scoring forms for professionals and non-professionals alike is crucial given the limited availability of research tools. Future research in both directions, focusing on detailed investigations of the evidence base and the development of user-friendly scoring tools, would be highly valuable. This underscores the overall need for continued research in the field of mHealth apps for depression and their evidence base. External professionals conducting thorough research on popular apps can help bridge the gap between developer claims and the actual evidence base.

References

Ahmad, A., Li, K., Feng, C., Asim, S. M., Yousif, A., & Ge, S. (2018). An Empirical Study of Investigating Mobile Applications Development Challenges. *IEEE Access*, *6*, 17711–17728. https://doi.org/10.1109/access.2018.2818724

Aitken, M., Nass, D., & IQVIA Institute. (2021). Digital Health Trends 2021: Innovation, Evidence, Regulation, and Adoption. *IQVIA Institute for Human Data Science*, 1–5.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Pub Inc.

American Psychological Association. (2021). *What is cognitive behavioral therapy?* APA. Retrieved January 7, 2023, from https://www.apa.org/ptsd-guideline/patients-and-families/cognitive-behavioral

Amick, H. R., Gartlehner, G., Gaynes, B. N., Forneris, C., Asher, G. N., Morgan, L. C., Coker-Schwimmer, E., Boland, E., Lux, L. J., Gaylord, S., Bann, C., Pierl, C. B., & Lohr, K. N. (2015). Comparative benefits and harms of second generation antidepressants and cognitive behavioral therapies in initial treatment of major depressive disorder: systematic review and meta-analysis. *BMJ*, h6019. https://doi.org/10.1136/bmj.h6019

Anthes, E. (2016). Pocket psychiatry: Mobile mental-health apps have exploded onto the market, but few have been thoroughly tested. *Nature*, 532(7597).

Apple Inc. (n.d.). *Submit your Apps today*. Apple Developer. Retrieved January 10, 2023, from https://developer.apple.com/app-store/submitting/

Azhar, S., Jackson, A., & Sattineni, A. (2015). Construction Apps: A Critical Review and Analysis. In *Proceedings of the . . . ISARC*. https://doi.org/10.22260/isarc2015/0008

Bamonti, P. M., Price, E., & Fiske, A. (2014). Depressive Symptoms and Suicide Risk in Older Adults: Value Placed on Autonomy as a Moderator for Men but Not Women. *Suicide and Life Threatening Behavior*, *44*(2), 188–199. https://doi.org/10.1111/sltb.12062

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Frontiers in Digital Health*, *4*. https://doi.org/10.3389/fdgth.2022.847991

BECK, A. T. (1964). Thinking and Depression. *Archives of General Psychiatry*, *10*(6), 561. https://doi.org/10.1001/archpsyc.1964.01720240015003

Beck, A. T. (1972). *Depression: Causes and Treatment*. University of Pennsylvania Press.

Beck, J. S. (2013). *Cognitive Behavior Therapy: Basics and Beyond* (2nd ed.). The Guilford Press.

Ben-Zeev, D., Brenner, C. J., Begale, M., Duffecy, J., Mohr, D. C., & Mueser, K. T. (2014). Feasibility, Acceptability, and Preliminary Efficacy of a Smartphone Intervention for Schizophrenia. *Schizophrenia Bulletin*, *40*(6), 1244–1253. https://doi.org/10.1093/schbul/sbu033

Bhat, A. A., Goud, B. R., Pradeep, J., Jayaram, G., Radhakrishnan, R., & Srinivasan, K. (2020). Can Mobile Health Improve Depression Treatment Access and Adherence Among Rural Indian Women? A Qualitative Study. *Culture, Medicine and Psychiatry*, *44*(4), 461–478. https://doi.org/10.1007/s11013-019-09664-3

Buijink, A. W. G., Visser, B. J., & Marshall, L. H. (2013). Medical apps for smartphones: lack of evidence undermines quality and safety. *BMJ Journals*, *18*(3), 90–92. https://doi.org/10.1136/eb-2012-100885

Carlo, A. D., Hosseini Ghomi, R., Renn, B. N., Strong, M. A., & Areán, P. A. (2020). Assessment of Real-World Use of Behavioral Health Mobile Applications by a Novel Stickiness Metric. *JAMA Network Open*, *3*(8), e2011978. https://doi.org/10.1001/jamanetworkopen.2020.11978

Ceci, L. (2022, November). *Mobile and Internet Apps*. Statista. Retrieved March 10, 2023, from https://www-statista-com./statistics/276623/number-of-apps-available-in-leading-app-stores/

Corrigan, P. W., Druss, B. G., & Perlick, D. A. (2014). The Impact of Mental Illness Stigma on Seeking and Participating in Mental Health Care. *Psychological Science in the Public Interest*, *15*(2), 37–70. https://doi.org/10.1177/1529100614531398

Dahmer, L. (2020). *Werde ich im Job benachteiligt, wenn ich eine Therapie mache?* ZEIT ONLINE. https://www.zeit.de/zett/2020-04/werde-ich-im-job-benachteiligt-wenn-ich-eine-therapie-mache?utm_referrer=https%3A%2F%2Fwww.google.com%2F

Darcy, A., Beaudette, A., Chiauzzi, E., Daniels, J., Goodwin, K., Mariano, T. Y., Wicks, P., & Robinson, A. (2022). Anatomy of a Woebot® (WB001): agent guided CBT for women with postpartum depression. *Expert Review of Medical Devices*, *19*(4), 287–301. https://doi.org/10.1080/17434440.2022.2075726

Daudt, H. M., Van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology*, *13*(1). https://doi.org/10.1186/1471-2288-13-48

Depression in adults: recognition and management | Guidance | NICE. (2009, October 28). https://www.nice.org.uk/guidance/cg90

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience*, *9*(10), 788–796. https://doi.org/10.1038/nrn2345

Dong, M., Zeng, L., Lu, L., Li, X., Ungvari, G. S., Ng, C. H., Chow, I. H., Zhang, L., Zhou, Y., & Xiang, Y. (2019). Prevalence of suicide attempt in individuals with major depressive disorder: a meta-analysis of observational surveys. *Psychological Medicine*, *49*(10), 1691–1704. https://doi.org/10.1017/s0033291718002301

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M., & Christensen, H. (2013). Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review. *Journal of Medical Internet Research*, *15*(11), e247. https://doi.org/10.2196/jmir.2791

Ferrari, A. J., Norman, R. E., Freedman, G., Baxter, A. J., Pirkis, J. E., Harris, M. G., Page, A., Carnahan, E., Degenhardt, L., Vos, T., & Whiteford, H. A. (2014). The Burden Attributable to Mental and Substance Use Disorders as Risk Factors for Suicide: Findings from the Global Burden of Disease Study 2010. *PLoS ONE*, *9*(4), e91936. https://doi.org/10.1371/journal.pone.0091936

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287–298. https://doi.org/10.1002/wps.20472

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Gartlehner, G., Hansen, R. A., Morgan, L. C., Thaler, K., Lux, L., Van Noord, M., Mager, U., Thieda, P., Gaynes, B. N., Wilkins, T., Strobelberger, M., Lloyd, S., Reichenpfader, U., & Lohr, K. N. (2011). Comparative Benefits and Harms of Second-Generation Antidepressants for Treating Major Depressive Disorder. *Annals of Internal Medicine*, *155*(11), 772. https://doi.org/10.7326/0003-4819-155-11-201112060-00009

Gautam, M., Tripathi, A., Deshmukh, D., & Gaur, M. (2020). Cognitive Behavioral Therapy for Depression. *Indian Journal of Psychiatry*, *62*(8), 223. https://doi.org/10.4103/psychiatry.indianjpsychiatry_772_19

GBD Results. (n.d.). Institute for Health Metrics and Evaluation. https://vizhub.healthdata.org/gbd-results/?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b

Gil, S., & Droit-Volet, S. (2009). Time perception, depression and sadness. *Behavioural Processes*, *80*(2), 169–176. https://doi.org/10.1016/j.beproc.2008.11.012

Gonzalez, V. M., & Dulin, P. L. (2015). Comparison of a smartphone app for alcohol use disorders with an Internet-based intervention plus bibliotherapy: A pilot study. *Journal of Consulting and Clinical Psychology*, *83*(2), 335–345. https://doi.org/10.1037/a0038620

Google Developers. (n.d.). *Upload your app to the Play Console*. Android Developers. Retrieved February 10, 2023, from https://developer.android.com/studio/publish/upload-bundle

Grossman, J. T., Frumkin, M. R., Rodebaugh, T. L., & Lenze, E. J. (2020). mHealth Assessment and Intervention of Depression and Anxiety in Older Adults. *Harvard Review of Psychiatry*, *28*(3), 203–214. https://doi.org/10.1097/hrp.0000000000000255

Grote, N. K., Zuckoff, A., Swartz, H., Bledsoe, S. E., & Geibel, S. (2007). Engaging Women Who Are Depressed and Economically Disadvantaged in Mental Health Treatment. *Social Work*, *52*(4), 295–308. https://doi.org/10.1093/sw/52.4.295

Grundy, Q., Wang, Z., & Bero, L. (2016). Challenges in Assessing Mobile Health App Quality. *American Journal of Preventive Medicine*, *51*(6), 1051–1059. https://doi.org/10.1016/j.amepre.2016.07.009

Grünzig, S. D., Baumeister, H., Bengel, J., Ebert, D. S., & Krämer, L. V. (2018). Effectiveness and acceptance of a web-based depression intervention during waiting time for outpatient psychotherapy: study protocol for a randomized controlled trial. *Trials*, *19*(1). https://doi.org/10.1186/s13063-018-2657-9

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Hähnel, A., Weigel, B., Hoyer, J., & Helbig, S. (2004). Wartezeit für Psychotherapiepatienten – und wie sie zu nutzen ist. *Verhaltenstherapie*, *14*(4), 294–302. https://doi.org/10.1159/000082839

Heffner, J. L., Vilardaga, R., Mercer, L. D., Kientz, J. A., & Bricker, J. B. (2014). Feature-level analysis of a novel smartphone application for smoking cessation. *The American Journal of Drug and Alcohol Abuse*, *41*(1), 68–73. https://doi.org/10.3109/00952990.2014.977486

Hoffmann, T., Bennett, S., & Del Mar, C. D. (2017). *Evidence-Based Practice Across the Health Professions*. Elsevier Gezondheidszorg.

Holden, M. T., & Lynch, P. M. (2004). Choosing the Appropriate Methodology: Understanding Research Philosophy. *The Marketing Review*, *4*(4), 397–409. https://doi.org/10.1362/1469347042772428

Huguet, A., Rao, S., McGrath, P. J., Wozney, L., Wheaton, M., Conrod, J., & Rozario, S. (2016). A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. *PLOS ONE*, *11*(5), e0154248. https://doi.org/10.1371/journal.pone.0154248

Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Institute Gated. (n.d.). https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/digital-health-trends-2021/iqvia-institute-digital-health-trends-2021.pdf?_=1628089218603

Kauer, S. D., Reid, S. C., Crooke, A. H. D., Khor, A., Hearps, S. J. C., Jorm, A. F., Sanci, L., & Patton, G. (2012). Self-monitoring Using Mobile Phones in the Early Stages of Adolescent Depression: Randomized Controlled Trial. *Journal of Medical Internet Research*, *14*(3), e67. https://doi.org/10.2196/jmir.1858

Kazdin, A. E. (2017). Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy*, *88*, 7–18. https://doi.org/10.1016/j.brat.2016.06.004

Kerber, A., Beintner, I., Burchert, S., & Knaevelsrud, C. (2021). Does app-based unguided self-management improve mental health literacy, patient empowerment and access to care for people with mental health impairments? Study protocol for a randomised controlled trial. *BMJ Open*, *11*(7), e049688. https://doi.org/10.1136/bmjopen-2021-049688

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder. *JAMA*, *289*(23), 3095. https://doi.org/10.1001/jama.289.23.3095

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617. https://doi.org/10.1001/archpsyc.62.6.617

Khder, M. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and Its Applications*, *13*(3), 145–168. https://doi.org/10.15849/ijasca.211128.11

Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V., Chapleau, M., Paquin, K., & Hofmann, S. G. (2013). Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review*, *33*(6), 763–771. https://doi.org/10.1016/j.cpr.2013.05.005

Kim, H. M., Xu, Y., & Wang, Y. (2022). Overcoming the Mental Health Stigma Through m-Health Apps: Results from the Healthy Minds Study. *Telemedicine and e-Health*, *28*(10), 1534–1540. https://doi.org/10.1089/tmj.2021.0418

Kitano, H. H. L. (1970). Mental Illness in four Cultures. *Journal of Social Psychology*, *80*(2), 121–134. https://doi.org/10.1080/00224545.1970.9712532

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization*, *82*(11), 858–866.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Lasalvia, A., Zoppei, S., Van Bortel, T., Bonetto, C., Cristofalo, D., Wahlbeck, K., Bacle, S. V., Van Audenhove, C., Van Weeghel, J., Reneses, B., Germanavicius, A., Economou, M., Lanfredi, M., Ando, S., Sartorius, N., Lopez-Ibor, J. J., & Thornicroft, G. (2013). Global pattern of experienced and anticipated discrimination reported by people with major depressive disorder: a cross-sectional survey. *The Lancet*, *381*(9860), 55–62. https://doi.org/10.1016/s0140-6736(12)61379-8

Lau, N., O'Daffer, A., Colt, S., Yi-Frazier, J. P., Palermo, T. M., McCauley, E., & Rosenberg, A. R. (2020). Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. *JMIR mHealth and uHealth*, *8*(5), e17798. https://doi.org/10.2196/17798

Le, T., Senaratne, H., McQuaid, M., & Tigwell, G. W. (2021). *Exploring a Multifaceted Framework to Support the Design of Mobile Apps for Self-Regulating Anxiety*. https://doi.org/10.1145/3411763.3451645

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, *5*(1). https://doi.org/10.1186/1748-5908-5-69

Li, L. S. E., Wong, L. L., & Yap, K. Y. L. (2021). Quality evaluation of stress, anxiety and depression apps for COVID-19. *Journal of Affective Disorders Reports*, *6*, 100255. https://doi.org/10.1016/j.jadr.2021.100255

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry*, *18*(3), 325–336. https://doi.org/10.1002/wps.20673

Liu, J., Ma, J., Wu, W., Chen, X., Huang, X., & Xu, L. (2017). Protecting Mobile Health Records in Cloud Computing. *ACM Transactions on Embedded Computing Systems*, *16*(2), 1–20. https://doi.org/10.1145/2983625

Loumidis, K. S., & Shropshire, J. (1997). Effects of waiting time on appointment attendance with clinical psychologists and length of treatment. *Irish Journal of Psychological Medicine*, *14*(2), 49–54. https://doi.org/10.1017/s0790966700002986

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & Schweizer, E. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, *6*(1), 95–106. https://doi.org/10.1093/clipsy.6.1.95

Lui, J. H. L., Marcus, D. K., & Barry, C. T. (2017). Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. *Professional Psychology: Research and Practice*, *48*(3), 199–210. https://doi.org/10.1037/pro0000122

Luxton, D. D., June, J. D., & Chalker, S. A. (2015). Mobile Health Technologies for Suicide Prevention: Feature Review and Recommendations for Use in Clinical Care. *Current Treatment Options in Psychiatry*, *2*(4), 349–362. https://doi.org/10.1007/s40501-015-0057-2

Mackey, R., Gleason, A., & Ciulla, R. (2022). A Novel Method for Evaluating Mobile Apps (App Rating Inventory): Development Study. *JMIR mHealth and uHealth*, *10*(4), e32643. https://doi.org/10.2196/32643

Malik, T., Ambrose, A. J., & Sinha, C. (2022). Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis. *JMIR Human Factors*, *9*(2), e35668. https://doi.org/10.2196/35668

Martinengo, L., Van Galen, L., Lum, E., Kowalski, M., Subramaniam, M., & Car, J. (2019). Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Medicine*, *17*(1). https://doi.org/10.1186/s12916-019-1461-z

Meheli, S., Sinha, C., & Kadaba, M. (2022). Understanding People With Chronic Pain Who Use a Cognitive Behavioral Therapy–Based Artificial Intelligence Mental Health App (Wysa): Mixed Methods Retrospective Observational Study. *JMIR Human Factors*, *9*(2), e35671. https://doi.org/10.2196/35671

Mehta, A., Niles, A. N., Vargas, J. H., Marafon, T., Couto, D. D., & Gross, J. J. (2021). Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study. *Journal of Medical Internet Research*, *23*(6), e26771. https://doi.org/10.2196/26771

Miralles, I., Granell, C., Díaz-Sanahuja, L., Van Woensel, W., Bretón-López, J., Mira, A., Castilla, D., & Casteleyn, S. (2020). Smartphone Apps for the Treatment of Mental Disorders: Systematic Review. *JMIR mHealth and uHealth*, *8*(4), e14897. https://doi.org/10.2196/14897

Mitchell, A. J., & Selmes, T. (2007). Why don't patients attend their appointments? Maintaining engagement with psychiatric services. *Advances in Psychiatric Treatment*, *13*(6), 423–434. https://doi.org/10.1192/apt.bp.106.003202

Moberg, C., Niles, A., & Beermann, D. (2019). Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. *Journal of Medical Internet Research*, *21*(6), e12556. https://doi.org/10.2196/12556

Moher, D., Stewart, L., & Shekelle, P. (2015). All in the Family: systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. *Systematic Reviews*, *4*(1). https://doi.org/10.1186/s13643-015-0163-7

Mohr, D. C., Ho, J., Duffecy, J., Baron, K. G., Lehman, K. A., Jin, L., & Reifler, D. (2010). Perceived barriers to psychological treatments and their relationship to depression. *Journal of Clinical Psychology*, n/a-n/a. https://doi.org/10.1002/jclp.20659

Mojica Ruiz, I. J., Nagappan, M., Adams, B., Berger, T., Dienst, S., & Hassan, A. E. (2016). Examining the Rating System Used in Mobile-App Stores. *IEEE Software*, *33*(6), 86–92. https://doi.org/10.1109/ms.2015.56

Monnier, D. (2020). Woebot : psychothérapie, suite et fin ? *Psychothérapies*, *Vol. 40*(2), 71–78. https://doi.org/10.3917/psys.202.0071

Morris, M. E., & Aguilera, A. (2012). Mobile, social, and wearable computing and the evolution of psychological practice. *Professional Psychology: Research and Practice*, *43*(6), 622–626. https://doi.org/10.1037/a0029041

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., & Deleeuw, W. (2010). Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness. *Journal of Medical Internet Research*, *12*(2), e10. https://doi.org/10.2196/jmir.1371

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0611-x

Ng, C. H. (1997). The Stigma of Mental Illness in Asian Cultures. *Australian and New Zealand Journal of Psychiatry*, *31*(3), 382–390. https://doi.org/10.3109/00048679709073848

Ng, C., How, C., & Ng, Y. (2017). Depression in primary care: assessing suicide risk. *Singapore Medical Journal*, *58*(2), 72–77. https://doi.org/10.11622/smedj.2017006

Nice Guidelines. (2022, June 29). *Depression in adults: treatment and management*. Nice. https://www.nice.org.uk/guidance/ng222

Nicholas, J. M., Shilton, K., Schueller, S. M., Gray, E., Kwasny, M. J., & Mohr, D. C. (2019). The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. *Jmir Mhealth and Uhealth*, *7*(4), e12578. https://doi.org/10.2196/12578

Nicol, G., Wang, R., Graham, S., Dodd, S., & Garbutt, J. (2022). Chatbot-Delivered Cognitive Behavioral Therapy in Adolescents With Depression and Anxiety During the COVID-19 Pandemic: Feasibility and Acceptability Study. *JMIR Formative Research*, *6*(11), e40242. https://doi.org/10.2196/40242

Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2009). Mental disorders, comorbidity and suicidal behavior: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry*, *15*(8), 868–876. https://doi.org/10.1038/mp.2009.29

Octoparse Data Inc. (2022). *Web Scraping Services & Software*. Octoparse. https://www.octoparse.com/product

Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2008). Health and the Mobile Phone. *American Journal of Preventive Medicine*, *35*(2), 177–181. https://doi.org/10.1016/j.amepre.2008.05.001

Peterson, J., Pearce, P. F., Ferguson, L. A., & Langford, C. A. (2017). Understanding scoping reviews. *Journal of the American Association of Nurse Practitioners*, *29*(1), 12–16. https://doi.org/10.1002/2327-6924.12380

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods*, *5*(4), 371–385. https://doi.org/10.1002/jrsm.1123

Pohl, M. (2023, February 1). *325000 mobile health apps available in 2017*. Research2guidance. https://research2guidance.com/325000-mobile-health-apps-available-in-2017/

Powell, A. C., Landman, A. B., & Bates, D. W. (2014). In Search of a Few Good Apps. *JAMA*, *311*(18), 1851. https://doi.org/10.1001/jama.2014.2564

Powell, A. C., Torous, J., Chan, S., Raynor, G. S., Shwarts, E., Shanahan, M., & Landman, A. B. (2016). Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. *JMIR mHealth and uHealth*, *4*(1), e15. https://doi.org/10.2196/mhealth.5176

Preziosa, A., Grassi, A., Gaggioli, A., & Riva, G. (2009). Therapeutic applications of the mobile phone. *British Journal of Guidance &Amp; Counselling*, *37*(3), 313–325. https://doi.org/10.1080/03069880902957031

Prochaska, J. J., Vogel, E. A., Chieng, A., Baiocchi, M., Maglalang, D. D., Pajarito, S., Weingardt, K. R., Darcy, A., & Robinson, A. (2021). A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug and Alcohol Dependence*, *227*, 108986. https://doi.org/10.1016/j.drugalcdep.2021.108986

Punton, G., Dodd, A. L., & McNeill, A. (2022). 'You're on the waiting list': An interpretive phenomenological analysis of young adults' experiences of waiting lists within mental health services in the UK. *PLOS ONE*, *17*(3), e0265542. https://doi.org/10.1371/journal.pone.0265542

Puyat, J. H., Kazanjian, A., Goldner, E. M., & Wong, H. (2016). How Often Do Individuals with Major Depression Receive Minimally Adequate Treatment? A Population-Based, Data Linkage Study. *The Canadian Journal of Psychiatry*, *61*(7), 394–404. https://doi.org/10.1177/0706743716640288

Radovic, A., Vona, P. L., Santostefano, A. M., Ciaravino, S., Miller, E., & Stein, B. D. (2016). Smartphone Applications for Mental Health. *Cyberpsychology, Behavior, and Social Networking*, *19*(7), 465–470. https://doi.org/10.1089/cyber.2015.0619

Rebedew, D. (2018, February 28). *Five Mobile Apps to Help Patients With Anxiety and Depression*. AAFP. https://www.aafp.org/pubs/fpm/issues/2018/0300/oa1.html

Roepke, A. M., Jaffee, S. R., Riffle, O. M., McGonigal, J., Broome, R., & Maxwell, B. (2015). Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. *Games for Health Journal*, *4*(3), 235–246. https://doi.org/10.1089/g4h.2014.0046

Roma, P., & Ragaglia, D. (2016). Revenue models, in-app purchase, and the app performance: Evidence from Apple's App Store and Google Play. *Electronic Commerce Research and Applications*, *17*, 173–190. https://doi.org/10.1016/j.elerap.2016.04.007

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A. A., Abdollahpour, I., Abdulkader, R. S., Abebe, H. T., Abebe, M., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., Abraha, H. N., . . . Bali, A. G. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1736–1788. https://doi.org/10.1016/s0140-6736(18)32203-7

Rumrill, P. D., Fitzgerald, S. M., & Merchant, W. R. (2010). Using scoping literature reviews as a means of understanding and interpreting existing literature. *Work*, *35*(3), 399–404. https://doi.org/10.3233/wor-2010-0998

Saraceno, B., Van Ommeren, M., Batniji, R., Cohen, A., Gureje, O., Mahoney, J., Sridhar, D., & Underhill, C. (2007). Barriers to improvement of mental health services in low-income and middle-income countries. *The Lancet*, *370*(9593), 1164–1174. https://doi.org/10.1016/s0140-6736(07)61263-x

Schnall, R., Rojas, M., Bakken, S., Brown, W. E., Jr, Carballo-Diéguez, A., Carry, M., Gelaude, D. J., Mosley, J. P., & Travers, J. L. (2016). A user-centered model for designing consumer mobile health (mHealth) applications (apps). *Journal of Biomedical Informatics*, *60*, 243–251. https://doi.org/10.1016/j.jbi.2016.02.002

Shen, N., Levitan, M. J., Johnson, A., Bender, J. L., Hamilton-Page, M., Jadad, A. a. R., & Wiljer, D. (2015). Finding a Depression App: A Review and Content Analysis of the Depression App Marketplace. *JMIR mHealth and uHealth*, *3*(1), e16. https://doi.org/10.2196/mhealth.3713

Sinha, C., Cheng, A. L., & Kadaba, M. (2022). Adherence and Engagement With a Cognitive Behavioral Therapy–Based Conversational Agent (Wysa for Chronic Pain) Among Adults With Chronic Pain: Survival Analysis. *JMIR Formative Research*, *6*(5), e37302. https://doi.org/10.2196/37302

Sirisuriya, D. S. (2015). A Comparative Study on Web Scraping. *Proceedings of 8th International Research Conference, KDU*.

Stawarz, K., Preist, C., Tallon, D., Wiles, N., & Coyle, D. (2018). User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. *Journal of Medical Internet Research*, *20*(6), e10120. https://doi.org/10.2196/10120

Su, L., & Anderson, P. L. (2022). User Behavior of a Publicly Available, Free-to-Use, Self-guided mHealth App for Depression: Observational Study in a Global Sample. *JMIR Formative Research*, *6*(10), e35538. https://doi.org/10.2196/35538

Swartz, H. A., Zuckoff, A., Grote, N. K., Spielvogle, H. N., Bledsoe, S. E., Shear, M. K., & Frank, E. (2007). Engaging depressed patients in psychotherapy: Integrating techniques from motivational interviewing and ethnographic interviewing to improve treatment participation. *Professional Psychology: Research and Practice*, *38*(4), 430–439. https://doi.org/10.1037/0735-7028.38.4.430

The Lancet. (2016). The health crisis of mental health stigma. *The Lancet*, *387*(10023), 1027. https://doi.org/10.1016/s0140-6736(16)00687-5

Thornicroft, G., Brohan, E., Kassam, A., & Lewis-Holmes, E. (2008). Reducing stigma and discrimination: candidate interventions. *International Journal of Mental Health Systems*, *2*(1), 3. https://doi.org/10.1186/1752-4458-2-3

Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L., Borges, G., Bruffaerts, R., Bunting, B., De Almeida, J. M. C., Florescu, S., De Girolamo, G., Gureje, O., Haro, J. M., He, Y., Hinkov, H., . . . Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. *British Journal of Psychiatry*, *210*(2), 119–124. https://doi.org/10.1192/bjp.bp.116.188078

Thornicroft, G., Mehta, N., Clement, S., Evans-Lacko, S., Doherty, M., Rose, D., Koschorke, M., Shidhaye, R., O'Reilly, C., & Henderson, C. (2016). Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *The Lancet*, *387*(10023), 1123–1132. https://doi.org/10.1016/s0140-6736(15)00298-6

Tomasic, M. (2022). *Major Depression (Unipolar Depression)*. Psycom. https://www.psycom.net/major-depressive-disorder

Torous, J., & Roberts, L. W. (2017). Needed Innovation in Digital Health and Smartphone Applications for Mental Health. *JAMA Psychiatry*, *74*(5), 437. https://doi.org/10.1001/jamapsychiatry.2017.0262

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/m18-0850

Uher, R., Payne, J. L., Pavlova, B., & Perlis, R. H. (2014). MAJOR DEPRESSIVE DISORDER IN DSM-5: IMPLICATIONS FOR CLINICAL PRACTICE AND RESEARCH OF

CHANGES FROM DSM-IV. *Depression and Anxiety*, *31*(6), 459–471. https://doi.org/10.1002/da.22217

Voderholzer, U., Beintner, I., Backes, B., Esguerra, E., & Hessler-Kaufmann, J. (2021). Implementing Videoconference CBT for Depression in Routine Outpatient Care: Outcome, Working Alliance, and Influence of Patients' Technology Commitment. *Verhaltenstherapie*, *31*(3), 238–247. https://doi.org/10.1159/000513643

Walker, D. H. (1997). Choosing an appropriate research methodology. *Construction Management and Economics*, *15*(2), 149–159. https://doi.org/10.1080/014461997373088

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-Month Use of Mental Health Services in the United States. *Archives of General Psychiatry*, *62*(6), 629. https://doi.org/10.1001/archpsyc.62.6.629

Wasil, A. R., Gillespie, S., Patel, R., Petre, A., Venturo-Conerly, K. E., Shingleton, R. M., Weisz, J. R., & DeRubeis, R. J. (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. *Journal of Consulting and Clinical Psychology*, *88*(11), 983–993. https://doi.org/10.1037/ccp0000604

Wasil, A. R., Gillespie, S., Shingleton, R., Wilks, C. R., & Weisz, J. R. (2020). Examining the Reach of Smartphone Apps for Depression and Anxiety. *American Journal of Psychiatry*, *177*(5), 464–465. https://doi.org/10.1176/appi.ajp.2019.19090905

Wasil, A. R., Palermo, E. H., Lorenzo-Luaces, L., & DeRubeis, R. J. (2022). Is There an App for That? A Review of Popular Apps for Depression, Anxiety, and Well-Being. *Cognitive and Behavioral Practice*, *29*(4), 883–901. https://doi.org/10.1016/j.cbpra.2021.07.001

Wasil, A. R., Patel, R., Cho, J. Y., Shingleton, R. M., Weisz, J. R., & DeRubeis, R. J. (2021). Smartphone apps for eating disorders: A systematic review of evidence-based content and application of user-adjusted analyses. *International Journal of Eating Disorders*, *54*(5), 690–700. https://doi.org/10.1002/eat.23478

Wasil, A. R., Venturo-Conerly, K. E., Shingleton, R. M., & Weisz, J. R. (2019). A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. *Behaviour Research and Therapy*, *123*, 103498. https://doi.org/10.1016/j.brat.2019.103498

WHO. (2021). *WHO*. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/depression

World Health Organization. (2015). *Preventing Suicide: A Global Imperative* (Illustrated).

Yousaf, O., Grunfeld, E. A., & Hunter, M. S. (2013). A systematic review of the factors associated with delays in medical and psychological help-seeking among men. *Health Psychology Review*, *9*(2), 264–276. https://doi.org/10.1080/17437199.2013.840954

Zhao, B. (2017). Web Scraping. *Springer International Publishing eBooks*, 1–3. https://doi.org/10.1007/978-3-319-32001-4_483-1

Zucker, L. (2020, June 2). *Versicherungen, Verbeamtung: Kann ich mir durch eine Psychotherapie meine Zukunft verbauen?* DER SPIEGEL, Hamburg, Germany. https://www.spiegel.de/psychologie/psychotherapie-welche-auswirkungen-auf-arbeitgeber-berufunsfaehigkeit-verbeamtung-a-1fe2bd28-dcb9-43bf-b406-3f133174d3f9

Appendix A

App Rating Inventory (ARI) checklist extended explanation for the Items Medical/Behavioural Focus, Direct/Indirect Evidence, and Theoretical Model.

This appendix consists of the longer version of the extended explanation for the Items Medical/Behavioural Focus, Direct/Indirect Evidence, and Theoretical Model.

Medical/Behavioural Focus. In item (a) a behavioural concern can be seen as a symptom of a behavioural change in a person, such as withdrawal from social contact, passivity, loss of daily routines (American Psychiatric Association, 2022; Uher et al., 2014). In the case of mental health issues, the medical concerns included emotional, cognitive, and physical symptoms, such as a depressed mood, diminished ability to think/ concentrate/ make decisions, or fatigue/ loss of energy (American Psychiatric Association, 2022). Concerns often indicated an underlying problem, such as in this case, depression. Since many applications only focused on a concern and mentioned the underlying problem itself, it was used to answer item (a).

This led to item (b), where it was checked if there was any form of testing or information about the criteria for a depression diagnosis available. There are a few test instruments to measure depression with the underlying criteria of the DSM-5, such as the PHQ-9 depression questionnaire (Kroenke & Spitzer, 2002). The PHQ-9 is a 9-item depression module from the full PHQ, which scores each of the 9 DSM-IV criteria that have already be mentioned in the introduction.

For item (c), clinical practice guidelines are recommendations for clinicians about the care of patients with specific conditions. For example, the American Psychiatric Association (APA) recently published practice guidelines that provide evidence-based advice for the assessment and treatment of mental disorders (American Psychiatric Association, 2022). APA's Clinical Practice Guideline recommends seven psychotherapy interventions for the treatment of depression, namely Behavioural therapy, Cognitive therapy, Cognitive-Behavorial therapy (CBT), Interpersonal therapy (IT), Mindfulness-based cognitive therapy (MBCT), Psychodynamic therapy, and Supportive Therapy. Also, the NICE guidelines give specific directions and evidence-based recommendations for the treatment of depression (Nice Guidelines, 2022). As introduced before, the guidelines also recommend different treatments for less severe depression (Nice Guidelines, 2022). The 11 available options include guided self-help, counselling, and SSRI antidepressants. Then, individual therapy options are recommended, such as CBT, behavioural activation, IT, and short-term psychodynamic psychotherapy Also, group therapy with focus on CBT, behavioural activation and mindfulness/meditation are mentioned.

Direct/Indirect Evidence. Formal research is described as using scientific methods, which findings should be expected to yield reliable knowledge which we can generalise (Holden & Lynch, 2004; Walker, 1997). Formal studies (item a) are conducted using a systematic approach, as well as scientific methods (Holden & Lynch, 2004). It is more scientific in discovering needed information or solving a problem and usually involves several components (Walker, 1997).

For a reference that shows that the app’s content is based on empirical literature (item b), it was not enough if there was information on an external website. The reference had to be already in the description and was, when available, checked for their correctness.

Theoretical Model. For item (a) it was not only important that the applications solely were based on one of the psychotherapy interventions, recommended by the APA or the guidelines by NICE. Rather, the content and tools had to be based on an empirically validated treatment model of the psychotherapy intervention. For example, CBT is based on the cognitive model of mental illness, initially developed by Beck (1964). In its simplest form, the cognitive model ‘hypothesizes that people’s emotions and behaviours are influenced by their perceptions of events. “It is not a situation in and of itself that determines what people feel but rather the way in which they construe a situation” (Beck, 1964; Beck, J.S., 2011).

Each intervention has an underlying treatment model (e.g., IT, MBCT, Mindfulness). To answer item (a), it was checked whether there was further information on how the intervention (e.g., CBT), if available, was implemented. Here, it had to be differentiated between the mention of an intervention and the implementation of further steps that build on an underlying model of the treatment. For example, it was not enough if only mindfulness was mentioned, it had to further be indicated how this treatment was implemented and with which tools/ functions. If available, it was checked if supporting tools could be found in the application (e.g., thought diary, meditation).