Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and healthcare professionals: development, feasibility, and potential effectiveness studies.

Gerard Anmella, Miriam Sanabra, Mireia Primé-Tous, Xavier Segú, Myriam Cavero, Ivette Morilla, Victoria Ruiz, Iria Grande, Ariadna Mas, Inés Martín-Villalba, Alejandro Caballo, Julia-Parisad Esteva, Arturo Rodríguez-Rey, Flavia Piazza, Francisco José Valdesoiro, Claudia Rodríguez-Torrella, Marta Espinosa, Carlota Sorroche, Giulia Virgili, Alicia Ruiz, Aleix Solanes, Joaquim Radua, María Antonieta Also, Elisenda Sant, Sandra Murgui, Mireia Sans-Corrales, Allan H.Young, Víctor Vicens, Jordi Blanch, Elsa Caballeria, Hugo López-Pelayo, Clara López, Victoria Olivé, Laura Pujol, Sebastiana Quesada, Brisa Solé, Anabel Martínez-Aran, Carla Torrent, Joana Guarch, Ricard Navinés, Andrea Murru, Giovanna Fico, Michele De prisco, Vicenzo Oliva, Casimiro Pio, Silvia Amoretti, María Fernández-Canceco, Marta Villegas, Eduard Vieta, Diego Hidalgo-Mazzei

Submitted to: Journal of Medical Internet Research on: October 07, 2022

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.
Table of Contents

Original Manuscript.................................................................................................................................................. 5
Supplementary Files............................................................................................................................................... 36
Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and healthcare professionals: development, feasibility, and potential effectiveness studies.

Gerard Anmella1* MD; Miriam Sanabra1 PsyD, PhD; Mireia Primé-Tous1 PsyD; Xavier Segú1 PsyD; Myriam Cavero1 MD, PhD; Ivette Morilla1 PsyD; Victoria Ruiz1 RNC; Iria Grande1 MD, PhD; Ariadna Mas1 PsyD; Inés Martín-Villalba1 PsyD; Alejandro Caballo1 PsyD; Julia-Parisad Esteva1 PsyD; Arturo Rodríguez-Rey1 PsyD; Flavia Piazza1; Francisco José Valdesoiro1 PsyD; Claudia Rodriguez-Torrella1 PsyD; Marta Espinosa1 PsyD; Carlota Sorroche1 PsyD; Giulia Virgili1 PsyD; Alícia Ruiz1 PsyD; Aleix Solanes2; Joaquim Radua2 MD, PhD; Maria Antonieta Also3 MD; Elisenda Sant3 MD; Sandra Murgui4 MD; Mireia Sans-Corrales4 MD, PhD; Allan H. Young5 PhD; Victor Vicens6 MD, PhD; Jordi Blanch1 MD, PhD; Elsa Caballeria1 MSc, PsyD; Hugo López-Pelayo1 MD, MSc; Clara López1 PhD, PsyD; Victoria Olivé7 MD; Laura Pujol7 PsyD; Sebastiana Quesada7 RNC; Brisa Solé7 PhD, PsyD; Anabel Martínez-Aran1 PhD, PsyD; Carla Torrent1 PhD, PsyD; Joana Guarch1 PhD, PsyD; Ricard Navínes1 MD, PhD; Andrea Murru1 MD, PhD; Giovanna Fico1 MD; Michele De prisco1 MD; Vicenzo Oliva1 MD; Casimiro Pio8; Silvia Amoretti1 PhD, PsyD; María Fernández-Canseco5; Marta Villegas5; Eduard Vieta1 MD, PhD; Diego Hidalgo-Mazzei1 MD, PhD

1Hospital Clínic de Barcelona, Department of Psychiatry and Psychology, Institute of Neuroscience Barcelona ES
2Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) Barcelona ES
3CAP Casanova, Consorci Atenció primaria de Salut Barcelona Esquerra (CAPSBE) Barcelona ES
4CAP Borrell, Consorci d’Atenció primaria de Salut Barcelona Esquerra (CAPSBE) Barcelona ES
5Centre for Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King’s College London GB
6Chief Medical Officer and co-founder of Abi Global Health Barcelona ES
7Hospital Clínic de Barcelona, Occupational Health Department barcelona ES
8Barcelona Supercomputing Center (BSC). Text Mining Technologies in the Health Domain Barcelona ES
9these authors contributed equally

Corresponding Author:
Diego Hidalgo-Mazzei MD, PhD
Hospital Clínico de Barcelona, Department of Psychiatry and Psychology, Institute of Neuroscience
C. de Villarroel, 170, 08036
Barcelona
ES

Abstract

Background: A significant proportion of people attending Primary Care (PC) have anxiety-depressive symptoms and work-related burnout and there is a lack of resources to attend them. The COVID-19 pandemic has worsened this problem, particularly affecting healthcare workers, and digital tools have been proposed as a workaround. We present the development, feasibility and effectiveness studies of chatbot (Vickybot) aimed at screening, monitoring, and reducing anxiety-depressive symptoms and work-related burnout in PC patients and healthcare workers.

Objective: Mitigate the growing problem of mental health problems in PC and among healthcare workers by developing a digital decision support platform combining machine-learning severity prediction models (phase 1) with a smartphone-based app for screening, monitoring and delivering evidence-based psychological interventions to people with anxiety and depressive symptoms, and work-related burnout (phase 2). Here we present the results of phase 2: the development, feasibility and effectiveness studies in PC patients and healthcare workers.

Methods: User-centered development strategies were adopted. Main functions included self-assessments, psychological modules, and emergency alerts. Healthy controls (HCs) tested Vickybot for reliability. (1) Simulation: HCs used Vickybot for 2 weeks to simulate different possible clinical situations and evaluated their experience. (3) Feasibility and effectiveness study: People consulting PC or healthcare workers with mental health problems were offered to use Vickybot for one month. Self-
assessments for anxiety (GAD-7) and depression (PHQ-9) symptoms, and work-related burnout (based on the Maslach Burnout Inventory) were administered at baseline and every two weeks. Feasibility was determined based on the combination of both subjective and objective user-engagement Indicators (UEIs). Effectiveness was measured using paired t-tests as the change in self-assessment scores.

**Results:** 40 HCs tested Vickybot simultaneously, and data was transmitted and registered reliably. (1) Simulation: 17 HCs (73% female; mean age=36.5±9.7) simulated different clinical situations. 98.8% of the expected modules were recommended according to each simulation. Suicidal alerts were correctly activated and received by the research team. (2) Feasibility and effectiveness study: 34 patients (15 from PC and 19 healthcare workers; 77% female; mean age=35.3±10.1) completed the first self-assessments, with 34 (100%) presenting anxiety symptoms, 32 (94%) depressive symptoms, and 22 (64.7%) work-related burnout. Nine (26.5%) patients completed the second self-assessments after 2-weeks of use. No significant differences were found for anxiety \[t(8) = 1.000, P = .34\] or depressive \[t(8) = .40, P = .70\] symptoms, but work-related burnout was significantly reduced \[t(8) = 2.87, P = .02\] between the means of the first and second self-assessments. There was a trend towards higher reduction in anxiety-depressive symptoms and work-related burnout with greater use of the chatbot. Three patients (8.8%) activated the suicide alert, and the research team intervened promptly with successful outcomes. Vickybot showed high subjective-UEIs, but low objective-UEIs (completion, adherence, compliance, and engagement). Feasibility was moderate.

**Conclusions:** The chatbot proved to be useful in screening the presence and severity of anxiety and depressive symptoms, in reducing work-related burnout, and in detecting suicidal risk. Subjective perceptions of use contrasted with low objective-use metrics. Our results are promising but suggest the need to adapt and enhance the smartphone-based solution in order to improve engagement. Consensus on how to report UEIs and validate digital solutions, especially for chatbots, are required.

(JMIR Preprints 07/10/2022:43293)
DOI: https://doi.org/10.2196/preprints.43293

**Preprint Settings**

1) Would you like to publish your submitted manuscript as preprint?
✔ Please make my preprint PDF available to anyone at any time (recommended).
   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
   Only make the preprint title and abstract visible.
   No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).
   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.
   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in "PubMed Now!" service, my accepted manuscript PDF will automatically be made openly available.
Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and healthcare professionals: development, feasibility, and potential effectiveness studies.

Gerard Anmella1,2,3,4,5†, Miriam Sanabra1,2,3,4,5†, Mireia Primé-Tous1, Xavier Seguí3, Myriam Cavero1,2,3,4,5, Ivette Morilla1,4, Iria Grande1,2,3,4,5, Victoria Ruiz3, Ariadna Mas1,2,3,4,5, Inés Martín-Villalba1, Alejandro Caballo1, Julia Parissad Esteva1, Arturo Rodríguez-Rey1, Flavia Piazza1, Francisco Jose Valdesoiro1, Claudia Rodriguez-Torrealla1, Marta Espinosa1, Giulia Virgili1, Carlota Sorroche1, Alicia Ruiz1, Aleix Solanes1,6, Joaquim Radua1,6,7,9, Antonieta Also Fontanet9, Elisenda Sant9, Sandra Murgui10, Mireia Sans-Corrales10, Allan H. Young11, Victor Vicens12, Jordi Blanch1,2,3,4,13, Elsa Caballera1,14, Hugo López-Pelayo1,14, Clara López1, Victoria Olivé15, Laura Pujol15, Sebastiana Quesada15, Brisa Solé1,2,3,4,5, Carla Torrent1,2,3,4,5, Anabel Martínez-Aran1,2,3,4,5, Joana Guarch1,2,3,4,5, Vicenzo Oliva1,2,3,4,5, Silvia Amoretti1,2,3,4,5, Casimiro Pío16, Maria Fernández-Canseco16, Marta Villegas16, Eduard Vieta1,2,3,4,5, Diego Hidalgo-Mazzei1,2,3,4,5†.

1. Hospital Clínic de Barcelona, Institute of Neuroscience, Department of Psychiatry and Psychology, Bipolar and Depressive Disorders Unit, Digital Innovation Group. c. Villarroel, 170, 08036 Barcelona, Catalonia, Spain.
2. Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), c. Villarroel, 170, 08036 Barcelona, Spain.
4. Department of Medicine, School of Medicine and Health Sciences, University of Barcelona (UB), c. Casanova, 143, 08036 Barcelona, Catalonia, Spain.
5. Institute of Neurosciences (UBNeuro).
6. Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, IDIBAPS, Barcelona, Spain.
7. Early Psychosis: Interventions & Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King’s College London, London, United Kingdom.
8. Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.
9. CAP Casanova, Consorci d’Atenció Primaria de Salut Barcelona Esquerra (CAPSBE), Barcelona, Catalonia, Spain.
10. CAP Borrell, Consorci d’Atenció Primaria de Salut Barcelona Esquerra (CAPSBE), Barcelona, Catalonia, Spain.
11. Centre for Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, United Kingdom.
12. Chief Medical Officer and co-founder of Abi Global Health
13. President of the European Association of Psychosomatic Medicine.
15. Hospital Clinic de Barcelona, Occupational Health Department, Barcelona, Spain.
16. Barcelona Supercomputing Center (BSC). Text Mining Technologies in the Health Domain.

† Shared first authorship.

*Corresponding author: Diego Hidalgo-Mazzei, MD, PhD, Digital Innovation Group, Bipolar and Depressive Disorders Unit, Department of Psychiatry and Psychology, Institute of Neuroscience, Hospital Clinic de Barcelona, 170 Villarroel st, 08036, Barcelona, Catalonia, Spain. University of Barcelona, IDIBAPS, CIBERSAM. Email: dahidalg@clinic.cat

Word count: Abstract 447, Main document (references not included) 5,694
Abstract

**Background:** A significant proportion of people attending Primary Care (PC) have anxiety-depressive symptoms and work-related burnout compounded by a lack of resources to meet their needs. The COVID-19 pandemic has exacerbated this problem and digital tools have been proposed as a solution. We present the development, feasibility, and potential effectiveness studies of Vickybot, a chatbot aimed at screening, monitoring, and reducing anxiety-depressive symptoms and work-related burnout in PC patients and healthcare workers.

**Methods:** User-centered development strategies were adopted. Main functions included self-assessments, psychological modules, and emergency alerts. Healthy controls (HCs) tested Vickybot for reliability. (1) Simulation: HCs used Vickybot for 2 weeks to simulate different clinical situations. (2) Feasibility and effectiveness study: People consulting PC or healthcare workers with mental health problems were offered to use Vickybot for one month. Self-assessments for anxiety (GAD-7) and depression (PHQ-9) symptoms, and work-related burnout (based on the Maslach Burnout Inventory) were administered at baseline and every two weeks. Feasibility was determined from both subjective and objective user-engagement Indicators (UEIs). Potential effectiveness was measured using paired t-tests or the Wilcoxon signed-rank test for the change in self-assessment scores.

**Results:** 40 HCs tested Vickybot simultaneously, and data was transmitted and registered reliably. (1) Simulation: 17 HCs (73% female; mean age=36.5±9.7) received 98.8% of the expected modules according to each simulation. Suicidal alerts were correctly received. (2) Feasibility and potential effectiveness study: 34 patients (15 from PC and 19 healthcare workers; 77% female; mean age=35.3±10.1) completed the first self-assessments, with 34 (100%) presenting anxiety symptoms, 32 (94%) depressive symptoms, and 22 (64.7%) work-related burnout. Nine (26.5%) patients completed the second self-assessments after two weeks of use. No significant differences were found between the scores of the first and second self-assessments for anxiety \([t(8) = 1.000, P = .34]\) or depressive \([t(8) = .40, P = .70]\) symptoms. However, work-related burnout scores were moderately reduced \((Z = -2.07, P = 0.038, r = .32)\). There was a non-significant trend towards higher reduction in anxiety-depressive symptoms and work-related burnout with greater use of the chatbot. Three patients (8.8%) activated the suicide alert, and the research team intervened promptly with successful outcomes. Vickybot showed high subjective-UEIs, but low objective-UEIs (completion, adherence, compliance, and engagement). Feasibility was moderate.

**Conclusions:** The chatbot was useful in screening for the presence and severity of anxiety and depressive symptoms, and detecting suicidal risk. Potential effectiveness was shown in reducing work-related burnout, but not anxiety or depressive symptoms. Subjective perceptions of use contrasted with low objective-use metrics. Our results are promising but suggest the need to adapt and enhance the smartphone-based solution in order to improve engagement. Consensus on how to report UEIs and validate digital solutions, especially for chatbots, are required.

**Keywords:** Primary Care; Healthcare workers; Depression; Anxiety; Symptom; Burnout; Digital; Smartphone; Chatbot; PRESTO.
Introduction

Between 30 and 50% of people attending Primary Care (PC) in Spain have mental health problems, with a large majority demonstrating mild-to-moderate anxiety and depressive symptoms, usually work-related [1,2]. Only 5% of these people will eventually require specialized mental healthcare [3]. Nevertheless, access to specialized mental health care can take up to 3 months due to the high number of referrals and the lack of resources [4–6], including personnel, consultation times, and lack of training of General Practitioners (GP) to approach mental health problems [7]. Considering this complex and problematic challenge, during the last decade, the Spanish National Health System has launched a Primary Care Mental Health Support Program (PCMHSP) composed of mental health specialists, including psychiatrists, psychologists, and nurses, who dedicate part of their workweek to PC [8]. Nonetheless, the care demand still far exceeds the resources currently available at PC, representing a global challenge across the international community [9,10]. Consequently, collaborative initiatives are being increasingly proposed to address this urgent issue [11].

The unresolved delay in receiving specialized attention added to the lack of resources in PC, may lead to an increased and unnecessary risk of exacerbation of symptoms, affecting the quality of life of patients and potentially increasing sick leaves, and, in the worst case scenario, irreparable consequences such as suicide [12]. Evidence shows that up to 30% of patients with depression reported not receiving any type of care in the previous year [13]. This situation has led to a significant increase in the unnecessary and premature prescription of psychotropic drugs (up to 50% of people attending PC) [14–17], as well as superfluous referrals to specialized care, and a lack of early detection of severe mental health problems.

Adding to this already challenging situation, the COVID-19 pandemic has caused a significant impact on the mental health of the general population, with increases of more than 25% both in anxiety and depressive disorders [18]. This has exacerbated the existing problems of mental healthcare in PC [19–21], and also impacted healthcare workers, especially those in frontline [22–25]. In particular, healthcare workers have reported increased rates of anxiety, depression, and posttraumatic stress disorder (PTSD) symptoms [26–28], with potentially severe mental health consequences [25,29,30]. This increase in mental health problems in healthcare workers is related to an overload of work and included symptoms such as mental and physical exhaustion, depersonalization, diminished sense of personal accomplishment, and reduced professional functioning, thus fulfilling criteria of work-related burnout [31,32], which is usually co-morbid with symptoms of anxiety and depression, such as fatigue, insomnia, and sadness [33,34]. In fact, work-related burnout is a major issue among the general population, particularly in healthcare workers [35], that can lead to increased sick leaves with a severe economic impact [2], or even major consequences, such as suicide.

This sudden worldwide increase in the demand for mental healthcare has proved challenging, with essential mental health services disrupted [36]. Digital psychological tools have been proposed as a potential solution [37–39]. In fact, some digital interventions through smartphone applications (apps) have proven effective in improving anxiety [40] and depressive symptoms [41], and work-related burnout [42–44], especially in PC settings [45,46], and also for people suffering from severe mental disorders such as bipolar disorder [47,48].

More recently, chatbots (virtual assistant with artificial intelligence), by directly implementing Natural Language Processing (NLP) techniques and powered by artificial intelligence, have been shown to increase engagement, usability, and effectiveness of these interventions by offering a more friendly and personalized approach to users [49,50]. In people suffering PTSD (e.g., war veterans), the anonymous nature of chatbots has also effectively increased the disclosure of traumatic events [51]. This may be of particular interest for healthcare workers who have been exposed to extreme and continuous pressure during the COVID-19 pandemic. Additionally, NLP techniques have proven useful both for the design of smartphone apps [52] and to identify differential patient profiles in order to tailor specific psychological interventions according to these patients.

characteristics [53]. Apps and chatbots allow for a continuous psychological intervention that adapts to the needs, schedules, and status of the patient in a dynamic way based on the progression of the patient, with the possibility to capture active or passive data related to symptoms’ evolution [54,55].

The PRESTO project [56], intends to mitigate the growing problem of mental health problems in PC and among healthcare workers by developing a digital decision support platform combining machine-learning severity prediction models (phase 1) with a smartphone-based app for screening, monitoring, and delivering evidence-based psychological interventions to reduce anxiety and depressive symptoms, and work-related burnout (phase 2).

In this work, we present the pilot results of phase 2: the development, and studies of feasibility and potential effectiveness in PC patients and healthcare workers. The primary aim was to assess the feasibility of the intervention. Secondary aims included assessing: (i) its use in screening and monitoring anxiety and depressive symptoms, and work-related burnout; (ii) the potential effectiveness of the intervention on reducing them; and (iii) the ability of the intervention to detect suicide risk.
Methods

The “Vickybot” intervention: Development and functions.

We developed a chatbot intended to screen and monitor anxiety and depressive symptoms and work-related burnout while providing evidence-based psychological interventions. User-centered development strategies were adopted with evidence of improved user retention and satisfaction with the end product and increased chances of engagement in long-term real-world clinical settings [57]. Considering the preferences and recommendations of patients and healthcare workers presented in the protocol [56], the research team outlined the priority aspects of the digital tool as well as the psychological contents to be included. The main functions included were (Figure 1):

1. Self-administered scales for screening and monitoring anxiety (General Anxiety Disorder-7 (GAD-7) [58,59]), depressive symptoms (Patient Health Questionnaire-9 (PHQ-9) [60,61]), and work-related burnout (evaluated through two questions based on the emotional exhaustion dimension items from the Maslach Burnout Inventory (MBI) [62,63]). Users were prompted to complete all self-assessments every two weeks.

2. Psychological modules according to the self-assessments. Different modules were recommended for each individual according to an algorithm that considered the severity of the scales’ scores (both global and item-based). There was a total of 12 short, sequential, and customizable modules: Ten modules for anxiety and depressive symptoms (1. Depressed mood, 2. Anxiety, 3. Apathy-anhedonia, 4. Depressive cognitions, 5. Suicidal thoughts, 6. Restlessness, 7. Decreased concentration, 8. Overthinking, 9. Irritability, and 10. Sleep disturbance), and two modules for the management of 11. Work-related stress and 12. Burnout. The psychological modules were based on eclectic therapy [64], mostly including cognitive-behavioral therapy (CBT) techniques [65,66], but also mindfulness, dialectical behavioral therapy (DBT), [67] and acceptance and commitment therapy (ACT) [68] strategies. Specific functions were psychoeducation and weekly objectives (based on CBT), mindfulness (based on the stress-reduction program [69]), self-soothing strategies (based on DBT), and metaphors (based on ACT). All modules started with the psychoeducation section to evaluate whether participants identified with the symptoms/situations presented, and afterwards the other functions were shown. Then, participants could navigate among the aforementioned functions and use those which they felt more comfortable with or found more useful. Modules 1 and 2 (Depressed mood and Anxiety) only included psychoeducation. Module 4 (Depressive cognitions) included an intervention for people experiencing a negative view of the self, the world, and/or the future. These cognitions include failure, rejection, loss, guilt, lack of self-confidence or self-esteem, hopelessness, helplessness, and lack of hope for improvement.

3. A chatbot system allowing users to access the former functions in a friendly way and also able to answer questions and detect emergency situations such as suicidal thoughts.

4. An emergency alert of suicidal thoughts. For users who scored on item 9 of the PHQ-9 (suicidal thoughts) or if the chatbot detected suicidal inputs using NLP, an alert was sent to the research team and the user was recommended to immediately go to the emergency department and provided with emergency resources (telephone number for health emergencies and nearby hospital locations).

5. Reminders for the “weekly objective” function from the modules. Users could choose an hour of the day to perform the directed activity and they had the opportunity to create an automatic notification. Other reminders were set for self-assessments monitoring every two weeks and to assess user-experience with the intervention after one month of use.

6. Audio register. After completing the self-assessments, participants were offered to register an audio file.
about their feelings or concerns for future voice analysis.

Figure 1. Scheme of the main functions of the intervention “Vickybot”.

Feasibility and potential effectiveness evaluation of the intervention

In order to consolidate the development and evaluate the feasibility and potential effectiveness of the intervention, the following phases were carried out sequentially:

Set-Up phase

Over the course of two weeks, 40 healthy controls (HCs) used Vickybot and were asked to test the different functionalities without specific instructions (e.g., responding to self-assessments, accessing the psychological modules, setting reminders, etc.). A technical test was performed to evaluate the stability and reliability of the data transmission between devices and the servers, the tolerance of calls-per-minute to the server were verified, and logs of bug were also collected. HCs were recruited via advertisement and psychiatric disorders were excluded through a semi-structured interview.

Simulation phase

A total of 17 HCs, all participants from the previous set-up phase, were instructed to use Vickybot for two more weeks and simulate different clinical situations indicated by the research team, in order to elicit specific functions from the chatbot. The clinical simulations were determined so as to identify specific clinical profiles (severe-moderate-mild anxiety and depressive symptoms, work-related burnout, isolated symptoms, suicidal thoughts, etc.), exposed in Table 1. The data recorded on the server were correlated with the patterns identified for each clinical situation assigned. Also, the treatment algorithm for customization of the psychological modules was tested and calibrated for each simulation. Finally, baseline patterns of each simulator were then compared to its own follow-up outcomes as a function of time to assess variability and longitudinal registers.
After the simulation phase, HCs were asked for their insights on the functions that they thought could be improved. Subjective User-Engagement Indicators (UEI) (acceptability, usability, and satisfaction) were assessed with the mHealth App Usability Questionnaire (MAUQ) [70,71] (Table S1), which includes 18 statements to determine the usability, acceptability, and satisfaction of mental health interventions.

Feasibility and potential effectiveness study: Participants, procedure, and measures

Inclusion criteria: (i) people referred to the PCMHSP from the GP or healthcare workers from Hospital Clínic de Barcelona referred to specialized mental healthcare from the Department of Occupational Health; (ii) between 18 and 65 years of age; (iii) have a compatible smartphone and sufficient digital skills to use a chatbot (access, chat-interaction, main functions use, etc.); (iv) accept and sign the informed consent for study participation. Exclusion criteria: (i) previous or present diagnosis of a severe mental disorder; (ii) risk of suicide detected during clinical consultations prior to the use of the chatbot.

Participation to the study was offered during a clinical consultation either with the GP (for people referred to the PCMHSP), or the Department of Occupational Health (for healthcare workers). Afterwards, our research team contacted potential participants and explained the details and conditions of the study and resolved doubts. If participants consented to participate, they received an SMS-invitation with access to a survey from Hospital Clínic de Barcelona, where they e-signed the informed consent for participation, facilitated sociodemographic and clinical data, and were given access to Vickybot (Figure 2). All participants used their personal phones and no economic compensation was provided for their participation.
After accessing Vickybot (Figure 1), the chatbot presented its functions and asked users to perform a baseline self-assessment of anxiety-depressive symptoms and work-related burnout. According to the first self-assessments results, a treatment plan consisting of psychological modules was generated for each user. The following month, users were prompted to interact with the chatbot and perform their treatment plan in a sequential way using personalized reminders. Users were also suggested to complete the self-assessments every two weeks by receiving a preconfigured reminder.

At the end of the feasibility and effectiveness study (one month after the first self-assessments), similarly to the simulation phase, users were prompted to assess their experience with the different functions of the chatbot. Subjective-UEI (acceptability, usability, and satisfaction) were assessed with the MAUQ (Table S1), and objective-UEI (completion, adherence, compliance, and engagement) were retrieved from the servers. Feasibility was assessed based on the combination of both subjective and objective UEIs.

**Statistical analyses**
Statistical analyses were conducted using IBM SPSS Statistics for Windows Version 28.0 (IBM Corp; Armonk, NY). The analyses of all subjects were considered up until their last interaction with the chatbot. Sociodemographic and clinical data of the sample, as well as details on the general use of the intervention and the specific functions used were evaluated using descriptive analyses. The Saphiro-Wilk test and visual inspection of Q-Q plots were used to assess whether continuous variables were normally distributed.

The main outcomes for potential effectiveness were the changes in anxiety-depressive symptoms and work-related burnout measured with self-assessment scales using paired t-tests or the Wilcoxon signed-rank test. Subgroup analyses were conducted to assess the influence of baseline symptoms and the effect of engagement on the effectiveness of the intervention. Also, to assess the relationship between use and effectiveness, we performed Pearson correlations on the number of times that modules were completed and the days of use with the change in the clinical scales’ score. The results of the MAUQ at the end of the first month of use (acceptability, usability, and satisfaction), as well as other objective-use UEIs (completion, adherence, compliance, and engagement), and feasibility were presented performing descriptive analyses and analyzed according to the definitions recently proposed by the International Society for Bipolar Disorders (ISBD) Big Data Task Force [72]. Statistical significance was set at two-tailed p≤0.05.

Ethical considerations

The PRESTO project was approved by the Clinical Research Ethics Committee (CEIC) at Hospital Clinic of Barcelona (HCB/2020/0735). All the data were collected and stored in encrypted and secure servers following the guidelines and standards of the 2018 European General Data Protection Regulation (GDPR), and were solely managed by the project researchers. The project is in line with the ethical standards by European experts on personalized interventions and precision psychiatry [73].
Results

Set-Up phase

A total of 40 HCs tested Vickybot simultaneously. They accessed different functions of the intervention, with all functions tested by at least 7 users. Data was transmitted and registered correctly and the servers proved solid with a rapidly increasing number of users.

Simulation phase

A total of 17 HCs (73% female; mean age=36.5±9.7) simulated different clinical situations while using Vickybot. Table 1 exposes the specific clinical simulations and the functions of the intervention to be assessed respectively.

| Table 1. Clinical simulations and functions of the intervention assessed |
|-----------------------------|---------------------------------|---------------------------------|
| User | Clinical situation to be simulated | Specific functions to be tested | Specific psychological modules to be activated |
| 3 | Severe depressive symptoms | | |
| 7 | Depressive symptoms with suicidal thoughts | Emergency alerts of suicidal thoughts (item-9 of the PHQ-9 and chatbot NLP-intent detection), Automatic alert to the research team | 5. Suicidal thoughts |
| 8 | Lack of anxiety-depressive symptoms and lack of work-related burnout | Tools and "null" function of the treatment algorithm | N/A |
| 9 | Insomnia (solely) | 10. Sleep disturbance |
| 10 | Recurring concerns (solely) | 8. Overthinking |
| 11 | Cognitive problems (solely) | 7. Decreased concentration |
| 12 | Irritability (solely) | 9. Irritability |
| 13 | Depressive cognitions such as guilt/lack of confidence or self-esteem (solely) | 4. Depressive cognitions |
| 14 | Apathy (solely) | 3. Apathy-anhedonia |
All simulators signed-up, performed the first self-assessments, and obtained a treatment plan with specific psychological modules. 98.8% of the expected modules were recommended to users according to the respective simulations (Table 1). Only one module was not offered in the respective treatment plan (user 4) due to a lower than expected score on the PHQ-9. Ten (59%) users were offered “extra” modules that were not expected in the simulation. This was mostly due to high scores on the clinical scales by simulators that were asked to report specific symptoms solely (users 9-16). The expected simulated severity on anxiety and depressive symptoms was achieved by 75% of users (1-4) with one scoring below the expected severity (user 4). 98.5% of the expected functions (Table 1) were tested by simulators according to each clinical situation with the expected outcomes. All simulators showed a progressive reduction on the scales’ scores accordingly to a clinical simulation of symptoms’ reduction. This was not the case for one user (user 17), who was instructed to simulate “erratic behavior”. Suicidal alerts were correctly activated by the expected users (4 and 7) and received by the research team. Nine (53%) users set up “weekly objectives”, with six (67%) of them rating their daily completion. Eleven (65%) users set reminders. Nine (53%) users registered an audio file after the self-assessments.

Eleven (64.7%) HCs responded the MAUQ after the simulation phase, with a mean score of 6.39±0.36 on a scale from 1 (strongly disagree) to 7 (strongly agree). 100% of simulators agreed on 14/18 statements of the MAUQ questionnaire. Most patients (72-90%) agreed on the remaining 4/18 statements. Vickybot showed high usability, satisfaction, and acceptability among the simulators. The most recurrently perceived weak-points to be improved were the lack of reminders and the user’s lack of capacity to configure/personalize them, the lack of incentives to answer the self-assessments or access the psychological modules, and not being understood by the chatbot many times.

Feasibility and effectiveness study

More than 300 patients (>150 from PC and >150 healthcare workers) were offered participation in the study by their GPs or the personnel from the Department of Occupational Health respectively. After being contacted and receiving instructions from our research team, 130 patients (55 from PC and 75 healthcare workers) received the SMS-invitation. A total of 34 patients (15 from PC and 19 healthcare workers; 26.2% of those who received the SMS-invitation) signed-up and performed the first self-assessments, with 34 (100%) patients presenting low-to-severe anxiety symptoms, 32 (94%) depressive symptoms, and 22 (64.7%) reporting work-related burnout.

Table 2 provides a summary of the sample characteristics. Most were female (76.5%). Mean age was 35.3±10.1. A minority used alcohol or cannabis (<6%). More than 30% had medical comorbidities or non-psychiatric drugs prescribed. Almost 90% had current psychiatric diagnoses, most commonly anxiety (50%), followed by adjustment and depressive disorders (20%), and more than 80% had psychiatric drugs prescribed.

**Table 2. Characteristics of the sample (N=34).**
Sociodemographic variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M±SD)</td>
<td>35.3±10.12</td>
</tr>
<tr>
<td>Sex: Females N (%)</td>
<td>26 (76.5)</td>
</tr>
</tbody>
</table>

Medical record

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical comorbidities: Yes N (%)</td>
<td>17 (47.1)</td>
</tr>
<tr>
<td>Non-psychiatric drugs prescribed: Yes N (%)</td>
<td>13 (38.2)</td>
</tr>
</tbody>
</table>

Current drug use

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannabis: Yes N (%)</td>
<td>2 (5.9)</td>
</tr>
<tr>
<td>Alcohol: Yes N (%)</td>
<td>1 (2.9)</td>
</tr>
</tbody>
</table>

Mental health record

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past psychiatric diagnosis N (%)</td>
<td>11 (29.4)</td>
</tr>
<tr>
<td>Current psychiatric diagnosis* N (%)</td>
<td>30 (88.2)</td>
</tr>
<tr>
<td>Anxiety disorders N (%)</td>
<td>18 (52.9)</td>
</tr>
<tr>
<td>Adjustment disorders N (%)</td>
<td>7 (20.6)</td>
</tr>
<tr>
<td>Depressive disorders N (%)</td>
<td>7 (20.6)</td>
</tr>
<tr>
<td>Personality disorders N (%)</td>
<td>1 (2.9)</td>
</tr>
<tr>
<td>Eating disorders N (%)</td>
<td>1 (2.9)</td>
</tr>
</tbody>
</table>

Psychiatric drugs prescribed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past psychiatric drugs prescribed N (%)</td>
<td>6 (17.6)</td>
</tr>
<tr>
<td>Current psychiatric drugs prescribed N (%)</td>
<td>28 (82.4)</td>
</tr>
<tr>
<td>Antidepressants N (%)</td>
<td>23 (67.6)</td>
</tr>
<tr>
<td>Benzodiazepines N (%)</td>
<td>10 (29.0)</td>
</tr>
<tr>
<td>Trazodone/Mirtazapine N (%)</td>
<td>6 (17.6)</td>
</tr>
<tr>
<td>Low-dose amitryptiline N (%)</td>
<td>2 (5.9)</td>
</tr>
<tr>
<td>Other N (%)</td>
<td>4 (11.8)</td>
</tr>
</tbody>
</table>

M: mean; SD: standard deviation; N: number.
*Diagnoses were of low-to-moderate severity.

Nineteen (55.9%) patients used Vickybot for several days, with a mean of 15.3±17.6 days of use. Nine (26.5%) patients completed the second self-assessments after 2-weeks of use, and two (5.9%) completed the third self-assessments after 4-weeks of use. Longitudinal self-assessment scores are presented in Table 3. The Shapiro-Wilk test and visual examination of the Q-Q plots showed that the GAD7 and PHQ9 scores had a normal distribution, whereas the distribution of the work-related burnout scores deviated from normality.

Table 3. Longitudinal self-assessment scores

<table>
<thead>
<tr>
<th></th>
<th>First/Baseline (N=34)</th>
<th>Second/2-weeks (N=9)</th>
<th>Third/4-weeks (N=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAD-7 (M±SD)</td>
<td>13.4±4.9</td>
<td>13.0±4.7</td>
<td>10.5±6.4</td>
</tr>
<tr>
<td>PHQ-9 (M±SD)</td>
<td>13.2±6.6</td>
<td>12.7±7.4</td>
<td>11.0±7.1</td>
</tr>
<tr>
<td>Work-related burnout (Md; IQR)</td>
<td>2.0; 2.0</td>
<td>0.0; 2.0</td>
<td>0.0; nc</td>
</tr>
</tbody>
</table>

IQR: interquartile range; M: mean; Md: Median; N: number; nc: not computable; SD: standard deviation.

Potential effectiveness of the intervention

No significant differences between the means of the first and second self-assessments were found for depressive [t(8) = .40, P = .70] or anxiety [t(8) = 1.00, P = .34] symptoms. This lack of significant differences was maintained for the third self-assessments. Work-related burnout scores were significantly lower after 2 weeks of using Vickybot (Z = -2.07, P = 0.038), with a moderate effect size (r = .32) (Table 3).

To assess the effect of baseline symptoms, we conducted sub-analyses on participants with mild-to-moderate depressive (PHQ-9=5-14) and anxiety (GAD-7=5-14) symptoms. Again, no significant differences were found for depressive [t(4) = .33, P = .75] or anxiety [t(4) = .00, P = 1.00] symptoms between the means of the first and second self-assessments.

To assess whether the effect of the intervention was influenced by the use of the chatbot (dose-related effects...
of the intervention), we correlated the number of times that modules were performed (active treatment) with the change in the clinical scales’ scores, and no significant associations were found neither for depressive \( r = -0.48; P = .16 \), anxiety \( r = 0.12; P = .76 \) symptoms, or work-related burnout \( r = 0.55; P = .13 \). We also correlated the “days using the chatbot” with the change in the clinical scales’ scores, and no significant associations were found for depressive \( r = -0.23; P = .58 \), anxiety \( r = -0.41; P = .36 \) symptoms, or work-related burnout \( r = 0.38; P = 0.39 \). Despite the non-significant results, there was a trend towards a reduction in anxiety and depressive symptoms with greater use of the chatbot. To assess the effect of “clinically significant doses” of the intervention, we conducted sub-analyses on patients with >=50% (moderate to high) engagement. No significant differences between the scores of the first and second self-assessments were found for depressive \( t(6) = -0.82, P = 0.44 \), anxiety symptoms, or work-related burnout \( Z = -1.89, P = 0.059 \).

**Functions used**

All patients obtained a treatment plan with specific psychological modules recommended based on the first self-assessments. Twenty-five (73.5%) patients did access their treatment plan and performed at least one psychological module. In total, 112 modules were performed by users (mean±SD=4.5±2.4). The most performed modules were 1. Depressed mood and 2. Anxiety (32 and 31 times respectively), followed by 11. Work-related stress and 12. Burnout (17 and 15 times, respectively). The modules corresponding to 4. Depressive cognitions, 5. Suicidal thoughts, 7. Decreased concentration, and 9. Irritability were not performed by any user. Within the modules, 100% of patients that accessed the modules performed psychoeducation, 40% the weekly objective, 24% sense-based therapy, 16% mindfulness, and 4% metaphors. Six (24%) patients set up reminders for their “weekly objective” and three (50%) of them rated their daily completion. Sixteen (64%) patients set reminders. Fourteen (41%) patients registered an audio file after the first self-assessments, and two registered new audio files after the second and third self-assessments.

Three patients (8.8%) activated the suicide alert. Vickybot recommended them to go to the psychiatric emergency department, provided the emergency phone numbers, and the research team contacted them by phone-call within 8 hours. All three patients responded and acknowledged the call. Before the call, one of the contacted patients had overdosed with benzodiazepines and suicidal intentions, therefore he was urgently referred to the emergency department with successful outcome. The other two contacted patients were referred to the outpatient specialized mental healthcare services, as their current suicidal risk was deemed low by the clinical evaluation.

**User-Engagement Indicators**

With regard to subjective-UEIs, eight (24%) patients responded the MAUQ (Table S1) one month after using Vickybot, with a mean score of 6.53±0.34 on a scale from 1 (strongly disagree) to 7 (strongly agree). All patients agreed on 15/18 statements of the MAUQ questionnaire. Most patients (80%) agreed on the remaining 3/18 statements. All patients agreed on the 12th MAUQ statement, evaluating “satisfaction” with the chatbot, and also on the 18th statement, evaluating “acceptability” of the intervention. Vickybot showed high usability, satisfaction, and acceptability by patients. The most usefully-perceived functions of the chatbot were (in order of rating from best to worst): “self-assessments”, “personalized treatment plan”, “detection of suicidal thoughts”, “contents of the psychological modules”, and the “weekly objective”.

Concerning objective-UEIs, eight (24%) patients completed the study at 1 month highlighting a low “completers rate”. Low-adherence was also observed as only two (5.9%) patients used the chatbot after 4-weeks. Patients used the chatbot actively for 42.5% days of those expected, thus showing a low “compliance”. Patients used the chatbot actively for 22.8% of the expected weeks (at least 5/7 days of each week), meaning that “engagement” was also considered as low. Overall, results showed that Vickybot had low completers’ rate, adherence, compliance, and engagement by patients.

Conversely, Vickybot showed high subjective-UEIs (usability, satisfaction, and acceptability), but low
objective-UEIs (completion, adherence, compliance, and engagement). Considering the contrast between subjective and objective UEIs (perceived usefulness and actual use respectively), Vickybot was deemed of moderate feasibility.
Discussion

Principal Results

This study details the development and evaluation of the feasibility and potential effectiveness of a chatbot for anxiety and depressive symptoms and work-related burnout both for PC and healthcare workers. This intervention forms part of a larger project which aims to generate a digital platform (PRESTO) combining machine-learning severity prediction models with a smartphone-based app for screening, monitoring, and delivering psychological interventions [56]. The smartphone-based intervention has been envisaged to cover some of the most concerning demands from the PC system, and also to provide personalized help to healthcare workers, as both have been affected by the COVID-19 pandemic. These multiple purposes require a stepwise and solid growth as well as the collaboration and coordination of specialists in different areas.

The chatbot was useful in screening anxiety and depressive symptoms, and work-related burnout in PC patients and healthcare workers using self-administered scales. Although there were no statistically significant reductions in anxiety and depressive symptoms on the follow-up self-assessments, there was a significant reduction of work-related burnout (present in more than 60% of our sample). Similarly to other digital solutions [74,75], Vickybot pointed towards potential effectiveness in reducing work-related burnout. Future research with increased statistical power are required to confirm these promising findings. There is a need to tackle this growing issue with measures from the occupational and healthcare systems, and digital tools could offer a potential solution to this highly prevalent problem. Finally, the chatbot allowed an accurate detection and prompt intervention of emergency situations (suicidal thoughts).

It is known that healthcare workers and people referred to specialized mental healthcare from PC are distinct in several aspects, including medical literacy and work-related conditions. Despite the differences between both groups, they also share several common aspects, such as a high prevalence of depressive and anxious symptoms. In fact, even the causes of work-related burnout are particular and highly specific (i.e., stress and lack of support at the workplace) its manifestations include mostly anxiety and depressive symptoms such as fatigue, insomnia, and sadness. Moreover, even healthcare workers share a common work environment, they are highly heterogeneous in their background and work situation, including people from medical staff, nursing staff, nursing assistants, administrative workers, management, cleaning staff, security, and logistical support. In fact, during the development phase, these considerations were debated and taken into account, but we found that despite the distinct characteristics between both groups, their problems and preferences for a mental health app were highly similar. In this study, sub-analyses between healthcare workers and primary care patients regarding UEIs and potential effectiveness outcomes were performed, but no differences were found. This reinforces the hypothesis that these two groups have more in common regarding mental health problems than specific differences.

It should be noted that severe anxiety and/or depressive symptoms (GAD-7/PHQ-9>=15) were observed in 43.7% and 42.4% patients, and 23.5% presented both severe anxiety and depression simultaneously. In addition to the reduced sample size and the low engagement, the high severity of symptoms in >40% of the sample may explain the lack of reduction in anxiety and depressive symptoms. Digital interventions have only been found effective in the reduction of mild-to-moderate depressive symptoms [41], but not for severe symptoms. Considering the characteristics of the target population is therefore of utmost importance when designing digital interventions. In our case, we managed to identify patients with severe anxiety and depression, which, it could be argued, are probably not suited for this kind of intervention as they require face-to-face visits with mental health specialists. In this sense, and according to the literature, chatbots may be useful screening tools for mental health symptoms by providing a “humanized” environment while also preserving anonymization [76,77].
Detecting high-severity patients is one of the core problems faced by the overloaded PC system. Nowadays the prioritization system has been shown as ineffective [78], and digital solutions able to screen for severity may be suitable to solve this problem. Vickybot was useful and accurate in detecting both patients with higher severity, in addition to people having suicidal thoughts, while also offering on-time tools for immediate help (i.e., emergency services), and alerting the research team for evaluation and proposing tailored interventions. Suicide is probably the most striking unsolved problem in mental health [79], and the COVID-19 consequences are likely to worsen this situation [80–83]. Suicidal behaviors are the most alarming emergency situation that can be encountered, and thus, tools allowing prompt detection and immediate interventions are crucial. Digital solutions may fulfill these requirements [84] with some interventions already showing positive evidence [85]. In fact, since this is a very sensitive issue, consensus statements on how to tackle suicide with digital technologies have recently come to light [86], and conclusive remarks regarding the future consensus on the applicability of digital solutions in mental health, in both research and clinical practice, are yet to come.

Some smartphone-based interventions have shown efficacy in outcomes such as well-being, quality of life, or perceived stress, which are considered more important than symptom reduction by patients [87]. Likewise interventions, such as psychoeducation, empower patients with strategies on how to cope with symptoms. As such, even when symptoms are not reduced, enhancing patients’ perceptions of control and self-management are key interventions to improve their quality of life [88–90]. Studies on digital health interventions should therefore focus on those parameters and not only on symptoms’ reduction in order to gain further insight into their effectiveness and real-world application.

Moreover, smartphone-based solutions usually include functions such as symptoms’ self-assessment, longitudinal monitoring, detecting emergency situations (e.g., suicidal thoughts), and specific tools based on psychological therapies such as CBT (e.g., behavioral therapy), or mindfulness. It is still not clear which of the former functions are more suitable for each patient profile in terms of their severity in anxiety and depression, specific symptoms’ presentation, or sociodemographic characteristics [91]. The sub-analyses and correlations performed in this study showed a non-significant trend towards a reduction in anxiety and depressive symptoms, and work-related burnout with greater use of the chatbot. However, the reduced sample size in this longitudinal assessment may have limited them. Future studies with increased statistical power are needed to shed light on the remaining unanswered questions.

Regarding UEIs, subjective perceptions of use (acceptability, usability, and satisfaction) were highly positive, both for HCs during the simulation phase and for patients during the feasibility and effectiveness study. Patients especially valued the ability to perform mental health self-assessments and the personalization of treatments, as well as the chatbot’s ability to detect urgent situations and offer immediate help. This contrasted with low objective-use metrics (completion, adherence, compliance, and engagement). This was probably due to the fact that the chatbot had a rigid and sequential structure that users had to follow (based on the usual clinical evaluation and treatment procedures), thus lacking flexibility of use, which is key for user retention. This was shown by the initial use of the chatbot on day 1, with all users completing the first self-assessments and accessing at least some modules of the treatment plan followed by a gradual reduction using the chatbot, leaving many modules incomplete. Furthermore, some modules have not been accessed by any of the users. Other factors that may have lowered engagement outcomes are probably the lack of incentives offered by the intervention (e.g., paucity of personalized reminders), or the limitations of the natural-language understanding of the chatbot, as stated in the comments of the simulation phase. A decrease in engagement over time on the use of digital tools has been reported across the existing literature [48,92,93]. For instance, studies with similar digital interventions report equivalent results on engagement, with most participants only using the app once or twice [94], or only 20% of participants providing the required information in the app during the study period, with a gradual loss of entries after day 1 [95]. This suggests that engagement is a widespread problem with digital interventions, and that it is key to develop strategies to mitigate it. In fact, interventions designed for brief interactions have shown higher engagement outcomes [96]. This suggests the need to adapt smartphone-based interventions to the ways people use these technologies (usually brief and concise) [41]. According to users’ comments, and
the evidence in the literature, we propose that future versions of the chatbot should aim to be more flexible by offering users alternative flows of use at all times, and also including gamification with user-incentives, adding personalization strategies on the psychological modules and reminders, and enhancing the chatbot’s interpretation and response capacities with NLP techniques.

Limitations

First, the limited sample size of the feasibility and effectiveness study, the low engagement with the intervention, and the high severity of anxiety and depressive symptoms of many of the patients included may have precluded the possibility of finding potential effectiveness of the intervention in reducing anxiety and depressive symptoms. Accordingly, further research with increased sample sizes, particularly of patients with mild-to-moderate anxiety and depressive symptoms, and longer follow-up are required to determine the effectiveness of the intervention.

Second, we did not assess PTSD symptoms, which are prevalent among healthcare workers [23], and other relevant outcomes such as well-being, quality of life, and perceived stress [87]. We suggest that future studies consider these aspects to achieve a holistic perception of the intervention by users.

Third, while Vickybot succeeded on cross-sectionally screening and establishing the severity of anxiety and depressive symptoms, and work-related burnout, it failed to provide follow-up (bi-weekly) information on the evolution of most of the sample. Notably, obtaining cross-sectional information on symptoms’ severity is key for a successful triage in PC settings, and also to identify emergency situations (e.g., suicidal thoughts), in which the intervention was useful. However, a key goal of the PRESTO project is to provide self-management tools for people with anxiety and depressive symptoms. Moreover, only 34 out of 130 patients (26.2%) that received the SMS-invitation signed-up and performed the first self-assessments. This represents a loss of almost 3 out of 4 potential participants, which is in line with other studies showing similar losses of potential participants at the initial phases [94]. We believe that this was due to the access and download difficulties, so that future versions of the chatbot should simplify this process as much as possible. To achieve this, in-app sign-up (e.g., ensuring all steps can be conducted within the app itself), as well as reducing the amount of data required from participants prior to access, may provide a solution to reduce the complexity of this process. Altogether, the accessibility to the chatbot as well as its functions need to be redesigned considering users preferences while also keeping in mind the relevance of monitoring patient evolution.

Finally, there is still a lack of consensus on how to assess the feasibility of smartphone-based interventions, especially chatbots [88,97]. Also, UEIs still need consistent and replicable standard definitions and valid measures with solid thresholds [98]. For the current study we applied the definitions and metrics recommended by the ISBD [72]. Although these guidelines are based on digital interventions for bipolar disorder, the majority of their concepts may be applicable to other types of interventions. In recent years, platforms and databases of mental health smartphone-based interventions have been developed [99,100] including information on the characteristics of the intervention, functions included, population for whom they are directed, evidence-based validation, professional support, and expert reviews among others. Future directions point towards international expert consensus providing homogeneous valid indications on digital solutions’ regulation [101] and clinical validation [102].

Future lines of the PRESTO project include (1) enhancing the chatbot NLP component, including natural-language understanding and generation capacities, (2) integrating the chatbot into a smartphone-application in order to facilitate accessibility and incorporate the aforementioned functions, (3) perform focus groups with patients and professionals to detect needs and points of improvement to increase engagement, and finally (4) integrating the app into a digital platform to be implemented in the PC system.
Conclusions

The chatbot was useful in screening the presence and severity of anxiety and depressive symptoms. Although anxiety and depressive symptoms were not significantly reduced, there were significant reductions in work-related burnout on the follow-up self-assessments, thus suggesting the potential effectiveness of Vickybot. Emergency situations were accurately identified and prompt interventions with successful outcomes provided. Subjective perceptions of use (acceptability, usability, and satisfaction) were high in contrast to low objective-use metrics (completion, adherence, compliance, and engagement), and feasibility was moderate. Our results are promising, but suggest the need to adapt and enhance the smartphone-based solution in order to improve engagement. Finally, consensus on how to report user-engagement indicators and validate digital solutions, especially for chatbots, are required.
Acknowledgements

We are grateful to all participants.

Gerard Anmella is supported by a Rio Hortega 2021 grant (CM21/00017) from the Spanish Ministry of Health financed by the Instituto de Salud Carlos III (ISCIII) and co-financed by the Fondo Social Europeo Plus (FSE+).

Miriam Sanabra is supported by a grant from the Baszucki Brain Research Fund.

Ariadna Mas is supported by the Agència de Gestió d’Ajudes Universitàries I de Investigació (AGAUR) - PANDÈMIES 2020 grant (PI047003) from the Generalitat de Catalunya.

Iria Grande thanks the support of the Spanish Ministry of Science and Innovation (MCIN) (PI19/00954) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación y el Fondos Europeos de la Unión Europea (FEDER, FSE, Next Generation EU/Plan de Recuperación Transformación y Resiliencia_PRTR ); the Instituto de Salud Carlos III; the CIBER of Mental Health (CIBERSAM); and the the Secretaría d’Universitats i Recerca del Departament d’Economia i Coneixement (2017 SGR 1365), CERCA Programme / Generalitat de Catalunya as well as the Fundació Clínic per la Recerca Biomèdica (Pons Bartran 2022-FRCP_B1_2022).

Allan Young’s independent research is funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Joaquim Radua is supported by a Miguel Servet II contract (CPII19/00009), funded by Instituto de Salud Carlos III (ISCIII) and co-funded by European Social Fund “Investing in your future”.

Carla Torrent has been supported through a ‘Miguel Servet’ postdoctoral contract (CPI14/00175), and a Miguel Servet II contract (CPII19/00018) and thanks the support of the Spanish Ministry of Innovation and Science (PI17/01066 and PI20/00344), funded by the Instituto de Salud Carlos III and co-financed by the European Union (FEDER) ‘Una manera de hacer Europa’.

Anabel Martinez-Aran thanks the support of the Spanish Ministry of Science and Innovation (PI18/00789, PI21/00787) integrated into the Plan Nacional de I+D+I and co-financed by ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER); the ISCIII; the CIBER of Mental Health (CIBERSAM); the Secretaría d’Universitats i Recerca del Departament d’Economia i Coneixement (2017 SGR 1365), the CERCA Programme, and the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00177.

Andrea Murru thanks the support of the Spanish Ministry of Science and Innovation (PI19/00672) integrated into the Plan Nacional de I+D+I and co-financed by the ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER).

Giovanna Fico is supported by a fellowship from “La Caixa” Foundation (ID 100010434) - fellowship code - LCF/BQ/DR21/11880019.

Silvia Amoretti has been supported by a Sara Borrell contract (CD20/00177), funded by Instituto de Salud Carlos III (ISCIII) and co-funded by European Social Fund “Investing in your future”.

Eduard Vieta thanks the support of the Spanish Ministry of Science, Innovation and Universities (PI15/00283, PI18/00805, PI19/00394, PI21/00787, CPII19/00009) integrated into the Plan Nacional de I+D+I and co-
financed by the Instituto de Salud Carlos III (ISCIII)-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER); the ISCIII; the CIBER of Mental Health (CIBERSAM); the Secretaria d’Universitats i Recerca del Departament d’Economia i Coneixement (2017 SGR 1365), and the CERCA Programme / Generalitat de Catalunya. We would like to thank the Departament de Salut de la Generalitat de Catalunya for the PERIS grant SLT006/17/00357.

Diego Hidalgo-Mazzee’s research is supported by a Juan Rodés JR18/00021 granted by the Instituto de Salud Carlos III (ISCIII).

**Conflicts of interest**

Gerard Anmella has received CME-related honoraria, or consulting fees from Janssen-Cilag, Lundbeck, Lundbeck/Otsuka, and Angelini, with no financial or other relationship relevant to the subject of this article.

Myriam Cavero has received grants and served as consultant, advisor or CME speaker for the following entities: Lundbeck, Esteve, Pfizer, with no financial or other relationship relevant to the subject of this article.

Iria Grande has received grants and served as consultant, advisor or CME speaker for the following identities: ADAMED, Angelini, Casen Recordati, Ferrer, Janssen Cilag, and Lundbeck, Lundbeck-Otsuka, Luye, SEI Healthcare outside the submitted work.

Roland Zahn has collaborations with EMIS PLC, Depsee Ltd, and EMOTRA, received honoraria from Lundbeck and Janssen and is a co-I for a Livanova-funded study, and provides private healthcare.

Allan Young has received honoraria for lectures and advisory boards for all major pharmaceutical companies with drugs used in affective and related disorders with no financial or other relationship relevant to the subject of this article.

Hugo López-Pelayo has received grants for training by Lundbeck with no financial or other relationship relevant to the subject of this article.

Andrea Murru has received grants and served as consultant, advisor or CME speaker for the following entities: Angelini, Idorsia, Lundbeck, Pfizer, Takeda, with no financial or other relationship relevant to the subject of this article.

Eduard Vieta has received research support from or served as consultant, adviser or speaker for AB-Biotics, Abbott, Abbvie, Adamed, Angelini, Biogen, Celon, Dainippon Sumitomo Pharma, Ferrer, Gedeon Richter, GH Research, Glaxo SmithKline, Janssen, Lundbeck, Organon, Otsuka, Rovi, Sage pharmaceuticals, Sanofi-Aventis, Shire, Sunovion, Takeda, and Viatris, out of the submitted work.

Diego Hidalgo-Mazzee has received CME-related honoraria and served as consultant for Abbott, Angelini, Ethypharm Digital Therapy and Janssen-Cilag with no financial or other relationship relevant to the subject of this article.
Funding

The PRESTO project has been funded by Fundació Clínic per a la Recerca Biomèdica through the Pons Bartran 2020 grant (PI046549).

The development of a version of the digital solution adapted to health workers is funded by the Spanish Foundation for Psychiatry and Mental Health (FEPSM), the Spanish Psychiatric Society (SEP), and the Spanish Society of Biological Psychiatry (SEPB) (PI046813).

The enhancement of the digital solution with Natural Language Processing techniques in a chatbot user-interface in collaboration with the text mining technologies in the health domain of the Barcelona Supercomputing Center (BSC) is funded by the Agència de Gestió d’Ajudes Universitàries I de Investigació (AGAUR) - PANDÈMIES 2020 grant (PI047003), from La Generalitat de Catalunya.

Author contributions

- Study planning, project conception and coordination – GA, DHM, JB
- Development of the digital solution – MP, XS, MS, AM, IM, AC, JPE, ME, MAR, FP, FJV, CRT, ME, GV, CS, AR, CL, VO, LP, SQ, BS, AMA, JG, RN, VV, CP, MFC, MV, GA, DHM
- Recruitment – MS, DHM, MC, IM, IG, VR, ES, SM, MS, AMA, RN, GA
- Data extraction – MS, GA
- Data analysis – JR, AS, DHM, GA
- Manuscript preparation – GA, MS, DHM
- Manuscript revision – all authors

Consent for publication

All participants were asked to provide written informed consent prior to their inclusion in the study.

Data Availability Statement

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


42. Roy A, Druker S, Hoge EA, Brewer JA. Physician Anxiety and Burnout: Symptom Correlates and a Prospective Pilot Study of App-Delivered Mindfulness Training. JMIR mHealth uHealth JMIR Mhealth Uhealth; 2020 Apr 1:8(4). doi: 10.2196/15608


44. Pace TWW, Zeiders KH, Cook SH, Sarsar ED, Hoyt LT, Mirin NL, Wood EP, Tatar R, Davidson RJ. Feasibility, Acceptability, and Preliminary Efficacy of an App-Based Meditation Intervention to Decrease Firefighter Psychological Distress and Burnout: A One-Group Pilot Study. JMIR Form Res JMIR Form Res; 2022 Jun 1;6(6). doi: 10.2196/34951


52. Petersen CL, Halter R, Kotz D, Loeb L, Cook S, Pidgeon D, Christensen BC, Batsis JA. Using Natural Language Processing and Sentiment Analysis to Augment Traditional User-Centered Design: Development and Usability Study. JMIR mHealth uHealth NLM (Medline); 2020 Aug 1;8(8):e16862. doi: 10.2196/16862

53. Fukuoka Y, Lindgren TG, Mintz YD, Hooper J, Aswani A. Applying natural language processing to understand motivational profiles for maintaining physical activity after a mobile app and accelerometer-based intervention: The mPED randomized controlled trial. JMIR mHealth uHealth JMIR Publications; 2018 Jun 1;6(6). doi: 10.2196/10042


64. Luis Martorell J. Psicoterapias: Escuelas y conceptos básicos, 2da Edición. 2014.


74. Michelsen C, Kjellgren A. The Effectiveness of Web-Based Psychotherapy to Treat and Prevent Burnout: Controlled Trial. JMIR Form Res JMIR Form Res; 2022 Aug 11;6(8):e39129. doi: 10.2196/39129


76. Hungerbuehler I, Daley K, Cavanagh K, Claro HG, Kapps M. Chatbot-Based Assessment of Employees’ Mental Health: Design Process and Pilot Implementation. JMIR Form Res JMIR Form Res; 2021 Apr 1;5(4). doi: 10.2196/21678


97. To QG, Green C, Vandelanotte C. Feasibility, Usability, and Effectiveness of a Machine Learning-Based Physical Activity Chatbot: Quasi-Experimental Study. JMIR mHealth uHealth JMIR Mhealth Uhealth; 2021 Nov 1;9(11). doi: 10.2196/28577

98. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: A review of

100. Mobile Health Index and Navigation Database, App Evaluation Resources from the Division of Digital Psychiatry at BIDMC. Available from: https://mindapps.org/ [accessed Feb 1, 2022]


Abbreviations

Primary Care (PC)
Healthy controls (HCs)
User-engagement Indicators (UEIs).
General Practitioners (GP)
Primary Care Mental Health Support Program (PCMHSP)
Posttraumatic stress disorder (PTSD)
Natural Language Processing (NLP)
General Anxiety Disorder-7 (GAD-7)
Patient Health Questionnaire-9 (PHQ-9)
Maslach Burnout Inventory (MBI)
Cognitive-behavioral therapy (CBT)
Dialectical behavioral therapy (DBT)
Acceptance and commitment therapy (ACT)
mHealth App Usability Questionnaire (MAUQ)
International Society for Bipolar Disorders (ISBD)
European General Data Protection Regulation (GDPR)
Supplementary Files