Psychological Medicine

cambridge.org/psm

Preventing depression using a smartphone app: a randomized controlled trial

Mark Deady[1], Nicholas Glozier[2], Rafael Calvo[3,4], David Johnston[1,5], Andrew Mackinnon[1], David Milne[6], Isabella Choi[2], Aimee Gayed[7], Dorian Peters[3,8], Richard Bryant[9], Helen Christensen[1] and Samuel B. Harvey[1]

[1]Black Dog Institute, Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia; [2]Central Clinical School, Brain and Mind Centre, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia; [3]School of Electrical and Information Engineering, University of Sydney, Sydney, NSW, Australia; [4]Dyson School of Design Engineering, Imperial College London, London, UK; [5]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK; [6]School of Systems Management and Leadership, Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW, Australia; [7]School of Psychiatry, Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia; [8]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK and [9]School of Psychology, University of New South Wales, Sydney, NSW, Australia

Original Article

Cite this article: Deady M *et al* (2022). Preventing depression using a smartphone app: a randomized controlled trial. *Psychological Medicine* **52**, 457–466. https://doi.org/10.1017/S0033291720002081

Received: 6 January 2020
Revised: 15 April 2020
Accepted: 27 May 2020
First published online: 6 July 2020

Key words:
Behavioural activation; depression; eHealth; mHealth; mindfulness; randomized control trial; smartphone; workplace mental health

Author for correspondence:
Mark Deady, E-mail: m.deady@unsw.edu.au

Abstract

Background. There is evidence that depression can be prevented; however, traditional approaches face significant scalability issues. Digital technologies provide a potential solution, although this has not been adequately tested. The aim of this study was to evaluate the effectiveness of a new smartphone app designed to reduce depression symptoms and subsequent incident depression amongst a large group of Australian workers.

Methods. A randomized controlled trial was conducted with follow-up assessments at 5 weeks and 3 and 12 months post-baseline. Participants were employed Australians reporting no clinically significant depression. The intervention group ($N = 1128$) was allocated to use *HeadGear*, a smartphone app which included a 30-day behavioural activation and mindfulness intervention. The attention-control group ($N = 1143$) used an app which included a 30-day mood monitoring component. The primary outcome was the level of depressive symptomatology (PHQ-9) at 3-month follow-up. Analyses were conducted within an intention-to-treat framework using mixed modelling.

Results. Those assigned to the *HeadGear* arm had fewer depressive symptoms over the course of the trial compared to those assigned to the control ($F_{3,734.7} = 2.98$, $p = 0.031$). Prevalence of depression over the 12-month period was 8.0% and 3.5% for controls and *HeadGear* recipients, respectively, with odds of depression caseness amongst the intervention group of 0.43 ($p = 0.001$, 95% CI 0.26–0.70).

Conclusions. This trial demonstrates that a smartphone app can reduce depression symptoms and potentially prevent incident depression caseness and such interventions may have a role in improving working population mental health. Some caution in interpretation is needed regarding the clinical significance due to small effect size and trial attrition.**Trial Registration** Australian and New Zealand Clinical Trials Registry (www.anzctr.org.au/) ACTRN12617000548336

Introduction

Over recent decades, mental health conditions have become the leading cause of long-term disability in most middle- and high-income countries (Harvey, Henderson, Lelliott, & Hotopf, 2009; Murray et al., 2012), with depression now the leading cause of disease burden worldwide (World Health Organization, 2017). To date, most of the effort to reduce the burden of these disorders has been targeted at ensuring treatment is given to those with manifest disorders. Although effective treatments are available, cost-effectiveness models suggest that even if optimal treatment were delivered to all of those in need, only 35–50% of the overall burden of these conditions would be alleviated (Andrews, Issakidis, Sanderson, Corry, & Lapsley, 2004). Consequently, effective intervention prior to the onset of full-diagnostic disorder is fundamental in responding to this problem (Cuijpers, Beekman, & Reynolds, 2012).

There is increasing evidence that prevention of mental disorders is possible (Cuijpers, Van Straten, & Smit, 2005). However, the costs associated with delivering most traditional face-to-face prevention programmes inhibit scalability and make population-wide roll-out of these programmes unfeasible (Solomon, Proudfoot, Clarke, & Christensen, 2015). A recent meta-analysis found emerging evidence that common psychotherapeutic techniques, delivered through eHealth and mHealth (healthcare practices supported by Internet or mobile phone technologies, respectively), may be effective in preventing common mental disorders in general

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

populations (Deady et al., 2017a). To date though, large-scale, incidence-focused studies are lacking (Richards et al., 2016). Equally as pressing is the need for high-quality studies incorporating attention-control conditions (Firth et al., 2017).

As the workplace is a dominant part of many adults' lives, it is a prime location for mental health prevention interventions (Tan et al., 2014). Furthermore, workplace issues are consistently reported as a major source of stress, which, in turn, has implications for a number of adverse psychological outcomes, including depression (Tennant, 2001). Whilst workplace stress and risk is not localized to any particular industry, those working in male-dominated industries (MDIs) have been found to have elevated rates of anxiety and mood disorders compared to other workers (Battams et al., 2014). The nature of work within these industries also highlights the limitations of conventional prevention programmes which are rarely accessible, flexible or tailored. Rapid growth in the areas of eHealth and mHealth represent new frontiers for delivering and targeting health interventions (Firth et al., 2017; Linardon, Cuijpers, Carlbring, Messer, & Fuller-Tyszkiewicz, 2019). The present intervention was developed with specific tailoring to these particular industries (Peters, Deady, Glozier, Harvey, & Calvo, 2018). Although other apps for depression exist (Firth et al., 2017), none have been specifically designed for use in a workplace context. Furthermore, non-traditional CBT-based apps are rare and previous work in this area focuses entirely on symptom reduction in an unwell population. This trial aimed to evaluate the effectiveness of a new smartphone app designed to reduce depression symptoms and subsequent incident depression amongst a large group of Australian workers. To our knowledge, this is the largest ever depression prevention trial and the largest ever trial of a mental health app.

Methods

Study design

A randomized controlled trial (RCT) was conducted in Australia with two parallel arms, comparing an app-based prevention intervention (*HeadGear*) and an attention-control app. The study was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000548336), with ethical approval from the University of New South Wales (UNSW) Human Research Ethics Committee (HC17021). The study was conducted in accordance with relevant guidelines as evidenced by the published study protocol (Deady et al., 2018). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Participants

Participants were working Australians, with MDIs selectively targeted. An MDI is defined as one in which ⩾70% of workers are male (Australian Bureau of Statistics, 2008). Two recruitment methods were employed: promotion of the study via a range of Australian industry partners; and social media (Facebook, Twitter) advertising and postings targeted towards MDIs. All onboarding was completed electronically, including online consent, whereby participants explicitly agreed to each of the consent declaration items and were able to print/email the information sheet.

Eligible participants were required to have a valid telephone number, own an Apple/Android-operating smartphone, be currently employed and reside in Australia. Participants were excluded if they did not have reliable access to the Internet, could not read English or failed to provide their phone number. Participants were also not included if they had substantial levels of depression symptoms at baseline, as indicated by a score above 14 on the Patient Health Questionnaire-9 (PHQ-9) or meeting provisional major depressive disorder (MDD) diagnosis using the PHQ-9 algorithm (Kroenke, Spitzer, & Williams, 2001).

Randomization and masking

Randomization occurred immediately following baseline assessment using automated procedures integrated into the trial management software. The randomization algorithm followed a block design with a block size of 10, to ensure an equal number of participants were assigned to each condition. As both the intervention and the attention-control condition had the same download and installation process, participants were blinded to their allocation.

Procedures

Interested individuals were directed to their respective app store via online advertisements or the study website. Upon downloading the app, participants provided informed consent and were screened. Participants in the intervention arm used *HeadGear*, a smartphone application-based intervention centred on behavioural activation (BA) and mindfulness (for full content, see study protocol) (Deady et al., 2018). The main component of the app is a 30-day intervention involving one 5–10 min 'challenge' per day. These 'challenges' feature evidence-based therapeutic techniques delivered using a variety of formats. At the commencement of the intervention, users complete a risk calculator that assesses risk for future common mental disorders and provides personalized feedback regarding this risk (Fernandez et al., 2017). The *HeadGear* app also includes a mood tracker, a toolbox of skills (which is filled as the intervention is completed) and support service helplines.

Participants in the attention-control arm used a smartphone application with an identical design as *HeadGear* which included the risk calculator and mood tracker; however, there was no 30-day intervention. To control for the attentional component of the intervention, the control condition encouraged users to track their mood daily over a 30-day period.

Assessments were completed at baseline, post-intervention (5 weeks after baseline), 3- and 12-month follow-up. The intervention and control applications monitored usage data including time spent in-app, number of logins and activity completion rates.

While baseline assessment occurred in the app, follow-up assessments occurred via an online survey accessed through an SMS web link. Participants received up to three phone-based reminders to complete each follow-up. On completion of each assessment, participants were entered into a draw for one of four $200 Visa gift cards.

The PHQ-9 was used to measure depression symptoms and was the primary outcome measure for this study (Spitzer, Kroenke, & Williams, 1999). The PHQ-9 is a reliable and valid nine-item measure of depression severity over the past 2 weeks, with each item scored on a four-point scale (Spitzer et al., 1999). Summing the nine items results in a score ranging from

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

0 (no depressive symptoms) to 27 (all symptoms occurring nearly daily). The criterion and construct validity of the PHQ-9 have previously been demonstrated (Spitzer et al., 1999). The tool also incorporates an algorithm for probable MDD diagnosis.

Anxiety was measured using the two-item Generalized Anxiety Disorder scale (GAD-2) (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007). Each item of the GAD-2 is scored on a four-point scale (total ranging from 0 to 6). Scale scores of $\geqslant 3$ are suggested as cut-off points between the normal range and probable cases of anxiety (Kroenke et al., 2007).

Resilience (the ability to adapt well in the face of adversity, trauma, tragedy and threats) was measured by the 10-item Connor–Davidson Resilience Scale (CD-RISC10), a self-report scale shown to have high internal consistency (Cronbach's $\alpha = 0.89$), construct validity and test–retest reliability in general population and in clinical settings (Connor & Davidson, 2003). Total scores range from 0 to 40 with higher scores corresponding to greater resilience. The CD-RISC10 has been shown to differentiate between individuals who function well after adversity and those who do not.

Well-being was assessed using the five-item WHO Wellbeing Index (WHO-5) (Bech, 2004). Raw scores range from 0 to 25, where 0 indicates the worst possible quality of life while a score of 25 represents the best possible quality of life. A score $\leqslant 13$ or an answer of 0 or 1 on any of the five items shows poor well-being. WHO-5 is a psychometrically sound measure of well-being with high internal consistency (Cronbach's $\alpha = 0.84$) and convergent associations with other measures of well-being (Krieger et al., 2014).

Work performance was measured using three items from the Health and Work Performance Questionnaire (HPQ) (Kessler et al., 2003) and an additional item pertaining to days absent in the last month. For the purposes of analysis, a composite measure for effective work days was constructed, by multiplying days present at work (absenteeism) by absolute work productivity score (presenteeism) as calculated by the HPQ, replicating previous work in the area (Wang et al., 2007).

Outcomes

The primary outcome measure of the study was the level of depressive symptomatology (as measured by the PHQ-9) across the 3-month follow-up period. Secondary outcomes included change in incident cases of depression over the follow-up period (using the PHQ-9 diagnostic algorithm), anxiety symptomatology, well-being, resilience and work performance, and 12-month outcomes.

Statistical analysis

As a universal preventative intervention with an attention-control, the size of the effect of the intervention was anticipated to be relatively small. Meta-analysis of previous trials of workplace prevention of depression has shown a small effect size of $d = 0.17$ (Tan et al., 2014). Power calculations were carried out using the R package simR (Green & MacLeod, 2016). A sample size of 1134 was needed to detect a difference of two points on the PHQ-9 (equivalent to a similar small effect size), at a significance level of 5% with a power of 80%. Using an attrition rate of 40%, an initial sample of 2100 was required.

Primary analyses were undertaken on an intent-to-treat basis, including all participants as randomized, regardless of treatment received or withdrawal from the study. Likelihood-based methods (mixed-model repeated measures) using IBM Statistical Package for the Social Sciences (SPSS) for Windows (release 23.0.0) (IBM Corp, 2015) were undertaken. *A priori* planned comparisons of change from baseline across the 3-month follow-up period were used to test the study hypotheses. Variables found to be substantially imbalanced between groups post-randomization were tentatively included in these models and retained if statistically significant and influential on outcomes. Mathematical transformation or categorization of raw scores was undertaken to meet distributional assumptions and address any violation of assumptions. For dichotomous outcomes such as depression caseness, a comparable generalized mixed modelling approach was used. Relative risk of depression was estimated based on the PHQ-9 diagnostic algorithm at both follow-up and the trial endpoint. Number needed-to-treat (NNT) was derived from the relative risk of new-onset depression at either follow-up time point (aggregated).

Little's test of the missing completely at random mechanism was conducted to determine the quality of missingness. Where this was significant, further analysis of patterns of missingness was undertaken to identify predictors of attrition which might indicate that data may meet the missing at random criteria.

All tests of treatment effects were conducted using a two-sided α level of 0.05 and 95% confidence intervals.

Results

A total of 5,578 participants were assessed for inclusion in the study. Of those screened, 3,500 failed to meet eligibility criteria, primarily because they had too many depressive symptoms to qualify for this trial of prevention and three did not complete baseline assessment. The remaining 2,275 eligible participants were randomized. Four participants later withdrew from the study. Thus, for the purpose of intent-to-treat analysis, there were 1,128 participants in the intervention group and 1,143 in the control group. The flow of participants through the study phases is shown in Fig. 1.

Table 1 presents the baseline characteristics and app usage of the study sample. The mean age was 40 years and participants were predominantly male (74.2%). A slightly greater proportion of females were randomized to intervention [compared to control; odds ratio (OR) = 1.21; corresponding to a small effect size]. Provisional adjustments were made to outcome analyses by incorporating gender into these models. No effects of this adjustment were found and consequently, these results are not reported. Primary outcome data were available post-intervention for 48% of the sample and for 46% at 3-month follow-up. Missingness was associated with intervention group assignment (OR = 1.65; $p < 0.001$; 95% CI 1.40–1.96) and younger age [OR(per year) = 0.98; $p < 0.001$; 95% CI 0.97–0.99]. Baseline depression level was unrelated to missingness.

Primary outcomes

The primary outcome was the level of depressive symptomatology. PHQ-9 scores (and model residuals) showed substantial non-normality; as such, transformations (square root and log) were applied to the data. Regardless of transformation type, very similar results were obtained. Log transformations are reported here, with raw data used for effect size.

There was a statistically significant group-by-time interaction for depression symptoms ($F_{2,1123.5} = 3.97$, $p = 0.019$ at primary

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

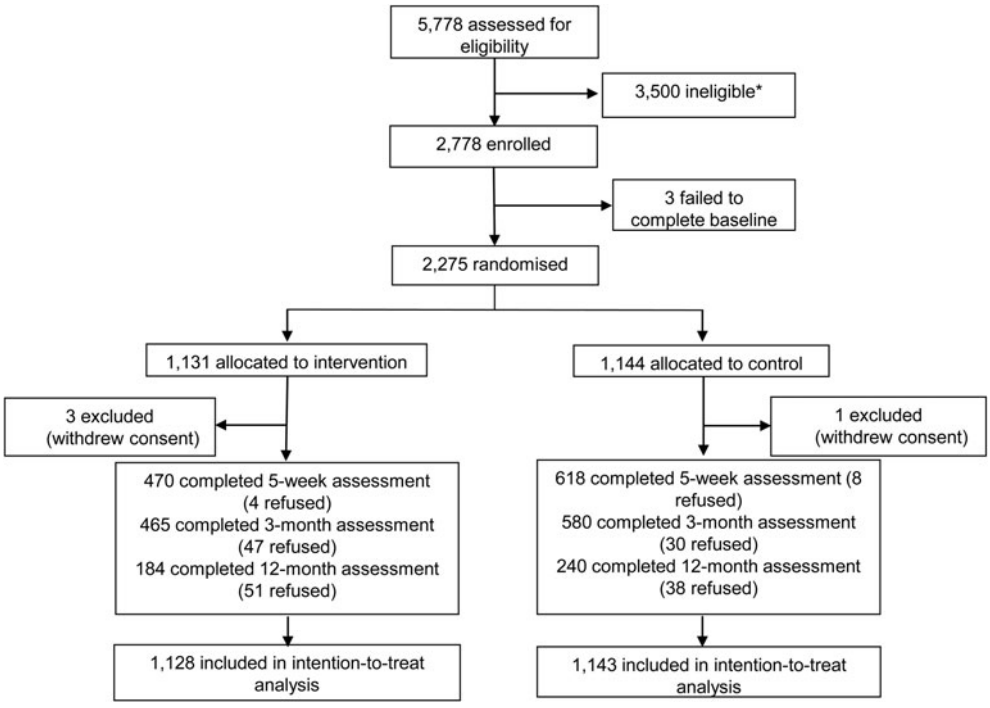


Fig. 1. Trial profile. *44 multiple registrations; 2,202 did not meet inclusion criteria (749 not employed, 29 not aged 18 or over, 1,424 depressed); 1,254 failed to provide contact information. Minor imbalance in groups occurred due to the removal of test users (erroneously included within the blocks).

3-month assessment, $F_{3,734.7} = 2.98$, $p = 0.031$ at 12-month; see Fig. 2). *HeadGear* was associated with a decline in mean PHQ-9 total score of 0.45 units more than the controls at post-intervention (Cohen's $d = 0.15$; 95% CI 0.03–0.27). At 3-month follow-up, the advantage of *HeadGear* was slightly lower, but remained statistically significant, at 0.41 units on the PHQ-9 ($d = 0.13$; 95% CI 0.01–0.27). Trial attrition was considerable at 12-month follow-up, with only 16.3% and 21.0% completing the assessment for intervention and control, respectively. At 12-month follow-up, group-by-time interaction for depression symptoms was not significantly different from baseline [$t_{(473.8)} = 1.50$, $p = 0.135$]. Comparisons of differential change between intervention arms by gender on individual occasions have different patterns of individual statistical significance. However, these differences are not statistically significant nor is the overall pattern of differential change between intervention arms (Supplementary file).

On average, users completed approximately a third of the core intervention component of the app, use generally tapered off but spiked at the completion of the entire programme (10.1%). When comparing outcome by 'challenge' completion, individuals in the intervention arm who completed few challenges (⩽7) were comparable to the control at follow-up, while those completing more challenges showed a more pronounced change, particularly those who completed 20 days or more [control *v.* 20–29 days, $t_{(1158.8)} = -2.73$, $p = 0.006$; control *v.* 30 days, $t_{(1158.8)} = -3.09$, $p = 0.002$].

Secondary outcomes

Cumulative incidence of depression

As a result of the inclusion criteria, no participants exceeded the cut-off for the established PHQ-9 algorithm for a provisional diagnosis of MDD at baseline. Post-intervention, 4.2% of participants in the control condition met the criterion for likely MDD, compared to 1.9% of those assigned to the intervention, leading to an OR of depression of 0.44 ($p = 0.038$; 95% CI 0.21–0.96). At 3-month follow-up, case rates for the control arm were stable at 5.5%, while rates in the intervention arm increased to 3.2% yielding an OR of 0.57 ($p = 0.079$; 95% CI 0.31–1.07; see Fig. 3). By the 12-month follow-up, depression prevalence for controls remained stable at 5.1%, while amongst participants in the *HeadGear* arm, the prevalence of depression remained significantly lower at 1.1%. This difference was associated with an OR of 0.21 ($p = 0.04$, 95% CI 0.05–0.96). Aggregated over all follow-up occasions, the prevalence of depression at any time was 8.0% and 3.5% for controls and *HeadGear* recipients, respectively. Odds of caseness for incident depression in the intervention group compared to controls were significant with an OR of 0.43 ($p = 0.001$, 95% CI 0.26–0.70) (Fig. 3). This equates to an NNT of 23 (17–44), suggesting one case of depression was avoided for every 23 users of *HeadGear*, compared to mood monitoring.

Work performance

A significant difference in mean change was observed for work performance (HPQ) at follow-up [$t_{(935.4)} = 3.03$, $p = 0.003$; see Table 2]. This was equivalent to an extra day of effective work each month amongst those using *HeadGear* compared to the control (1.19; 95% CI 1.10–1.29).

Resilience, well-being, anxiety

Other secondary outcomes are presented in Table 2. Significant differences in mean change were observed for resilience (CD-RISC10) at follow-up. For well-being (WHO-5), the difference in mean change was significant at post-intervention but not follow-up. While there was a decline from baseline to post-

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

Table 1. Selected demographics and app usage by intervention arm

	Intervention ($n = 1128$)	Control ($n = 1143$)	All ($n = 2271$)
Age (years; S.D.; range)	40.20 (10.88;18–78)	40.32 (10.38;18–68)	40.26 (10.63; 18–78)
Sex			
Male[a]	817 (72.4%)	869 (76.0%)	1686 (74.2%)
Female	311 (27.6%)	274 (24.0%)	585 (25.8%)
Industry			
Agriculture, forestry and fishing	23 (2.0%)	19 (1.7%)	42 (1.8%)
Wholesale trade	21 (1.9%)	28 (2.4%)	49 (2.2%)
Utilities, maintenance and manual trade	77 (6.8%)	86 (7.5%)	163 (7.2%)
Transport	238 (21.1%)	222 (19.4%)	460 (20.3%)
Mining	29 (2.6%)	37 (3.2%)	66 (2.9%)
Construction	67 (5.9%)	69 (6.0%)	136 (6.0%)
Safety	116 (10.3%)	113 (9.9%)	229 (10.1%)
Manufacturing	13 (1.2%)	8 (0.7%)	21 (0.9%)
Retail trade	33 (2.9%)	27 (2.4%)	60 (2.6%)
Accommodation and food services	23 (2.0)	28 (2.4%)	51 (2.2%)
Information media and telecommunications	13 (1.2%)	11 (1.0%)	24 (1.1%)
Financial and insurance services	35 (3.1%)	35 (3.1%)	70 (3.1%)
Public administration	36 (3.2%)	40 (3.5%)	76 (3.3%)
Professional, scientific and technical services	111 (9.8%)	128 (11.2%)	239 (10.5%)
Administration and support services	22 (2.0%)	21 (1.8%)	43 (1.9%)
Education and training	89 (7.9%)	92 (8.0%)	181 (8.0%)
Healthcare and social assistance	120 (10.6%)	132 (11.5%)	252 (11.1%)
Arts and recreation services	32 (2.8%)	23 (2.0%)	55 (2.4%)
Electricity, gas, water and waste (and other) services	12 (1.1%)	8 (0.7%)	20 (0.9%)
Unknown	18 (1.6%)	16 (1.4%)	34 (1.5%)
Male-dominated workforce			
MDI[b]	588 (52.1%)	586 (51.3%)	1174 (51.7%)
Non-MDI	522 (46.3%)	542 (47.4%)	1064 (46.9%)
Missing	18 (1.6%)	15 (1.3%)	33 (1.5%)
Role			
General employee	688 (61.0%)	719 (62.9%)	1407 (62.2%)
Manager	313 (27.7%)	289 (25.3%)	602 (26.5%)
Senior manager	122 (10.8%)	131 (11.5%)	253 (11.2%)
Missing	5 (0.4%)	4 (0.3%)	9 (0.4%)
Risk category[c]			
Low (⩽19.47%)	297 (26.3%)	269 (23.5%)	566 (24.9%)
Moderate (19.48–48.09%)	277 (24.6%)	290 (25.4%)	567 (25.0%)
High (⩾48.10%)	301 (26.7%)	265 (23.2%)	566 (24.9%)
Missing	253 (22.4%)	319 (27.9%)	572 (25.2%)
PHQ-9	6.9 (3.6)	6.9 (3.5)	6.9 (3.5)
Minimal or no symptoms (<5)	311 (27.6%)	325 (28.4%)	636 (28.0%)
Mild symptoms (5–9)	536 (47.5%)	510 (44.6%)	1046 (46.06%)
Moderate symptoms (10–14)	281 (24.9%)	308 (26.9%)	589 (25.94%)

(*Continued*)

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

Table 1. (*Continued.*)

	Intervention ($n = 1128$)	Control ($n = 1143$)	All ($n = 2271$)
App usage			
Total activity time (mins)[d]	52.29 (53.82)	22.64 (19.53)	37.37 (43.00)
Number of logins[d]	11.52 (13.17)	14.29 (18.95)	12.92 (16.29)
Number of moods logged	13.72 (16.82)	12.87 (18.46)	13.29 (17.66)
Challenge days completed	9.09 (10.08)	NA	NA

MDI, male-dominated industry; PHQ-9, Patient Health Questionnaire-9.
Data are *n* (%) or mean (s.d.).
[a]$p < 0.05$.
[b]Agriculture/forestry/fishing, manufacturing, wholesale trade, mining, construction, other manual trade, transport/postal/warehousing, first responder/defence/security.
[c]Category based on tertile split.
[d]$p < 0.001$.

intervention for anxiety symptoms (GAD-2), means in each arm were very similar. Exploratory analyses using a cut-off score of 3 to dichotomize outcomes suggested that the proportion of participants with clinically significant GAD was lower in the intervention group, but this result was not statistically significant. Significant differences in mean change were observed for well-being (WHO-5) and work performance (HPQ) at 12-month follow-up (Table 2).

Discussion

To our knowledge, this study is the first RCT to examine the effectiveness of a smartphone app in preventing depression and the largest mental health prevention trial undertaken in a working population. Overall, this study demonstrates that a smartphone app can reduce depression symptoms, may in fact prevent incident depression caseness, and improve work performance. Given the rising cost of mental illness across the developed world (Harvey et al., 2009), and the increasing awareness of the role workplaces can play in the development of mental illness, these findings make an important contribution to how new technology can be used to reduce the burden of common mental disorders.

Smartphones have become a pervasive feature of modern life and have transformed many aspects of our society, including healthcare. Published surveys have shown that over three-quarters of adults would be interested in using an app to monitor and manage their mood or mental health (Proudfoot et al., 2010). Developers have responded rapidly to this interest, with over 10 000 mental health-related applications now available to download, a number which increases daily (Torous & Roberts, 2017). However, science has failed to keep pace with these changes and almost all these apps are untested. This study provides evidence, for the first time, that a smartphone app may indeed have preventative potential regarding depression caseness. Using a diagnostic algorithm for depression, *HeadGear* was shown to be protective compared to simple mood monitoring, with rates of depression onset in the control group more than twice that

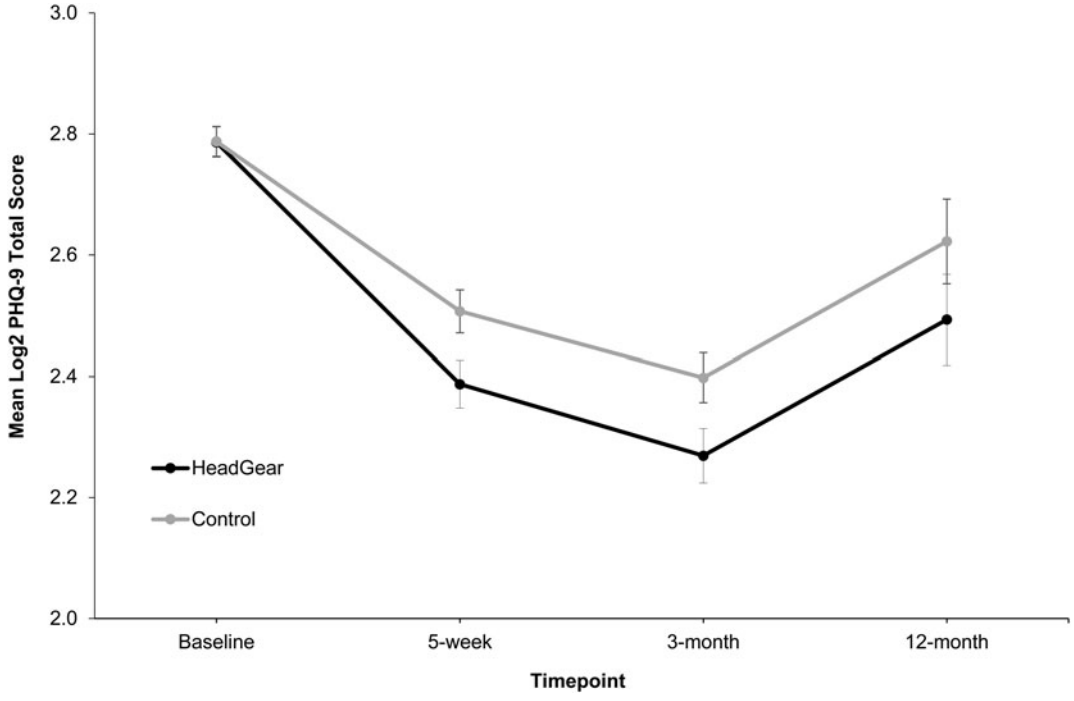


Fig. 2. Mean log2 PHQ-9 total scores by intervention arm at each occasion.

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

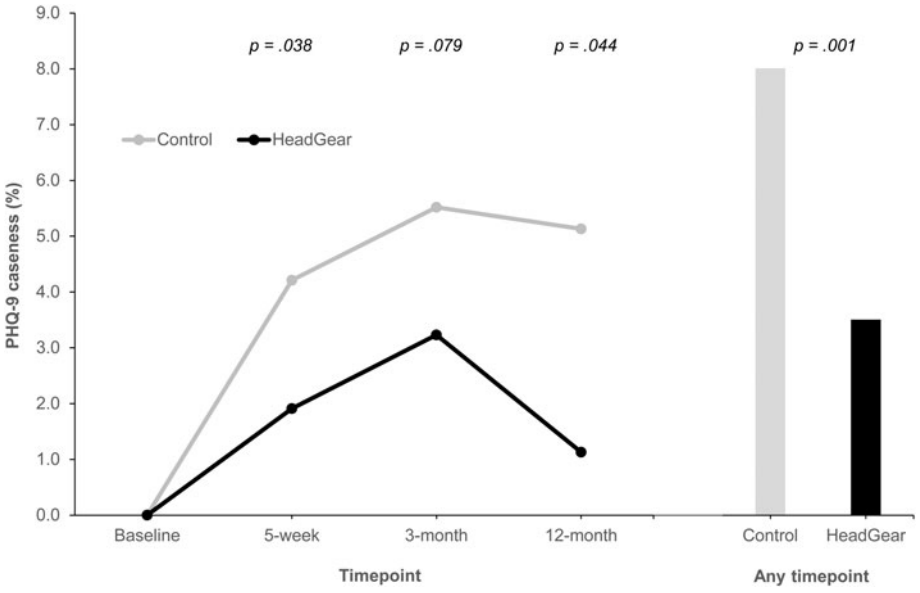


Fig. 3. PHQ-9 caseness by intervention arm at each occasion of measurement.

of workers randomized to use *HeadGear*. This indicates that one incident case of depression could be prevented for every 23 workers using *HeadGear*, which is comparable to the rates reported in a recent meta-analysis of depression prevention programmes (van Zoonen et al., 2014) and superior to many studies of CBT prevention interventions. This suggests the preventative impact of *HeadGear* may be substantial compared to the current best practice interventions. However, it is worth noting the effect size in the present study was small and may have implications for broader utility.

Many eHealth studies aim to intervene at a subclinical level, and subsequently, there is evidence of greater effectiveness within these groups (Firth et al., 2017; Heber et al., 2017). While the effect sizes for universal preventative and low-level symptom reduction interventions such as this one are small in a relative sense (Tan et al., 2014), the effect at a population level is considerable (Ahern, Jones, Bakshis, & Galea, 2008). The potential to move whole distributions away from the threshold of diagnosis is substantial, especially where dissemination can be so extensive, feasible and economical (Rose, 1994). The change observed in the control condition in the current trial may additionally be due to the use of the risk calculator and mood tracker. The latter may have influenced participants' mood states and the former may have led some control participants to seek assistance. Although the effect size reported compares favourably with recent meta-analyses in the area of mHealth CBT-related apps (Rathbone, Clarry, & Prescott, 2017) and workplace depression prevention (Tan et al., 2014), the question about how clinically meaningful small changes in symptom scores remains to be seen and tempers conclusions around impact. Furthermore, these findings should be interpreted with some caution due to the rates of attrition; more work is required to both improve programme adherence and trial dropout in this regard.

Table 2. Contrasts of differences in change from baseline to post-intervention and follow-up between conditions (EMM)

Effect	Mean difference (95% CI)	*t*	df	*p* value
WHO-5 post-intervention	0.63 (0.15–1.11)	2.55	1143.04	0.011
WHO-5 3-month follow-up	0.63 (−0.03 to 1.29)	1.85	822.98	0.065
WHO-5 12-month follow-up	1.13 (0.26–1.99)	2.56	502.70	0.011[a]
HPQ post-intervention	−0.23 (−1.32 to 0.86)	-0.42	1166.14	0.677
HPQ 3-month follow-up	1.04 (0.19–1.89)	2.41	661.62	0.016[a]
HPQ 12-month follow-up	1.73 (0.60–2.86)	3.01	464.10	0.003[a]
CD-RISC10 post-intervention	0.30 (−0.19 to 0.79)	1.20	1098.51	0.230
CD-RISC10 3-month follow-up	0.72 (0.04–1.39)	2.09	798.85	0.037[a]
CD-RISC10 12-month follow-up	0.74 (−0.22 to 1.69)	1.52	467.94	0.130
GAD-2 post-intervention	−0.38 (−0.28 to 0.09)	−1.02	1286.5	0.321
GAD-2 3-month follow-up	−0.09 (−0.29 to 0.09)	−0.99	845.35	0.113
GAD-2 12-month follow-up	−0.19 (−0.43 to 0.05)	−0.70	350.00	0.487

WHO-5, HO Wellbeing Index; HPQ, Health and Work Performance Questionnaire; CD-RISC10, Connor–Davidson Resilience Scale; GAD-2, Generalized Anxiety Disorder scale.
[a]$p < 0.05$

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

The economic case for depression prevention has often been promoted and is an essential component of decisions regarding how different public health interventions should be prioritized and funded (Cuijpers et al., 2012). For this reason, it is essential to test whether interventions can not only prevent symptom development, but also improve functioning. *HeadGear* was associated with significantly improved workplace performance and resilience at 3-month follow-up. Interestingly, this effect appeared to accumulate after the intervention had been delivered and was most apparent at follow-up. These findings indicate that the benefits associated with the app were broader than depression symptom or incidence reduction and that clear functional gains were also evident. The improvement in work performance amongst those randomized to receive *HeadGear* was equivalent to an extra day of effective work each month, which is a considerable functional and economic benefit. Although the app was not intended to specifically improve anxiety outcomes, the use of the GAD-2 (rather than a more rigorous measure) may have concealed these effects.

Our findings provide compelling evidence of the utility of mHealth tools, even in a clinically well population. A crucial element of this study [lacking in many eHealth trials (Firth et al., 2017)] was the use of an attention-control. Although controls spent less time engaging with this 'lite' version of the app, this is not an uncommon occurrence. Control conditions in this field tend to be significantly less intensive than interventions which often require users to engage with significant learning and rehearsal (Firth et al., 2017). In fact, there is a tendency for purposely inert attention-controls, while the content delivered here has been shown to have therapeutic benefit (Kauer et al., 2012). This study also provides valuable evidence to support the use of specific therapeutic components in depression prevention. To date, much of the research examining prevention and treatment interventions for depression has focused on CBT-based techniques (Deady et al., 2017a). Recent findings suggest BA (Deady et al., 2017b) and mindfulness-based (Stratton et al., 2017) techniques may be preferred by some types of workers. This trial adds to growing evidence of the effectiveness of such intervention techniques broadly (Richards et al., 2016), and the mobile delivery of such programmes.

Despite the unique strengths of the trial, some limitations are present. Firstly, there was substantial trial attrition at all time-points, particularly at 12 months, which limits interpretation of findings, especially long-term impact. A key factor which is unique to this trial, and may be implicated in this dropout rate, was the onboarding process. In order to streamline the user experience, completion of the full screening, consent and baseline assessment occurred within the app. This may have led to the study recruiting a less research-engaged (though perhaps more real-world) sample. Therefore, some caution is required in the conclusions reached as those completing follow-up may differ from those who do not in ways not measured. Similarly, on average, most users completed a third of the intervention days. Although not uncommon (Richards & Richardson, 2012), better outcomes were found with greater challenge completion, and further work should explore means to enhance user engagement to this end. Furthermore, the potential for regression to the mean cannot be discounted without a waitlist control. In attempting to remove barriers to recruitment, no diagnostic interview occurred at baseline, as such the lack of an initial diagnosis of depression could not be verified. Dual criteria of exclusion [>14 on the PHQ-9 (sensitivity = 68%, specificity = 95%) (Kroenke et al., 2001) and/or meeting the algorithm's criteria] were used in an attempt to attain a baseline sample free of MDD (Van Hooff et al., 2014). However, this approach may have resulted in some mild cases being misclassified as sub-syndromal, and thus a potential conflating of the concepts of prevention and very early treatment. Additionally, it was unfeasible to conduct full assessments on participants at follow-up, but this does preclude the ability to make definitive statements regarding true disorder prevention. Finally, although the gender imbalance present was intended due to the design of the intervention, as depression is more prevalent in females, the gender imbalance is a study limitation particularly with respect to the generalizability of our findings. However, concerns regarding this limitation are tempered by supplementary analyses that revealed gender had neither substantial confounding nor moderating effects. This is more that needs to be done to explore the utility of mHealth for mental health prevention.

Overall, the new mental health smartphone application, *HeadGear*, was associated with a significant reduction in depressive symptoms compared to an attention-control condition. The findings also suggest *HeadGear* may have preventative capacity for incident caseness of depression. Although a small effect size, combined with high rates of attrition, particularly at long-term follow-up, necessitate some caution is required in interpretation. Nevertheless, the study indicates that smartphone apps may have the capacity to empower individuals and enhance the way working adults manage their mental health. By utilizing new technologies, many of the feasibility obstacles encountered with more traditional interventions in this area are offset. The fusion of evidence-based techniques and new technology means we are now able to contemplate depression prevention initiatives operating at scale, meaning the prevention gains seen with chronic physical illnesses may finally be realized in mental health.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720002081

Acknowledgements. This study was developed in partnership with beyondblue with donations from the Movember Foundation. RC is funded by an Australian Research Council Future Fellowship FT140100824. SH and MD are supported by funding from the iCare Foundation and NSW Health. In addition to our funders, we would like to acknowledge the support of industry partners, including, but not limited to: NSW Fire and Rescue, Aurizon, Victoria SES, UNSW Facilities Management, EML, Transport NSW, Sydney Trains, Australia Post, Dairy NSW, Johns Lyng Group. We would also like to acknowledge Daniel Collins in helping prepare this manuscript for publication. We would like to acknowledge and thank all the study participants who have helped shape the content of the intervention.

Authorship. MD, DJ and SBH had full access to all the data in the study and take responsibility for the integrity of the data. AM had full access to the cleaned dataset and takes responsibility for the accuracy of the data analysis. MD had a primary role in the conceptualization, write up and editing of this manuscript. DJ, SBH and IC had a role in the conceptualization, write up and editing of this manuscript. NG had a role in the conceptualization and editing of this manuscript. DM had a role in the technical development of the intervention and editing of this manuscript. AM had a primary role in the statistical conceptualization and editing of this manuscript. RC, AG, RB and HC had a role in the write up and editing of this manuscript. All authors have read and approved the final manuscript.

Financial support. HC, AM, DJ, MD, SH, NG, RC, DM, DP, AG, IC, report grants from beyondblue (with donations from the Movember Foundation), during the conduct of the study. MD and SH report grants from iCare Foundation and NSW Health, during the conduct of the study. In addition,

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

MD, IC, SH, DM, RC, DP, NG have a patent IP from app design pending. RB has no interests to disclose. All researchers have remained independent from the funders in the completion and submission of this work. MD, DJ, IC, NG, DM, RAC, DP and SH were involved in the development of the *HeadGear* application. The IP is jointly owned by MD, IC, NG, DM, RC, DP and SH; however, the authors do not currently receive any financial gain from this IP. There are no other conflicts of interest or competing interests to declare. The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

References

Ahern, J., Jones, M. R., Bakshis, E., & Galea, S. (2008). Revisiting rose: Comparing the benefits and costs of population-wide and targeted interventions. *The Milbank Quarterly*, *86*(4), 581–600. doi: 10.1111/j.1468-0009.2008.00535.x

Andrews, G., Issakidis, C., Sanderson, K., Corry, J., & Lapsley, H. (2004). Utilising survey data to inform public policy: Comparison of the cost-effectiveness of treatment of ten mental disorders. *British Journal of Psychiatry*, *184*, 526–533.

Australian Bureau of Statistics. (2008). Australian and New Zealand standard industrial classification. Retrieved from Canberra: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1292.0.55.0052008?OpenDocument.

Battams, S., Roche, A. M., Fischer, J. A., Lee, N. K., Cameron, J., & Kostadinov, V. (2014). Workplace risk factors for anxiety and depression in male-dominated industries: A systematic review. *Health Psychology and Behavioral Medicine*, *2*(1), 983–1008. doi: 10.1080/21642850.2014.954579

Bech, P. (2004). Measuring the dimensions of psychological general well-being by the WHO-5. *Quality of Life Newsletter*, *32*, 15–16. doi: 10.1159/000289073

Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, *18*(2), 76–82. doi: 10.1002/da.10113

Cuijpers, P., Beekman, A. T. F., & Reynolds, C. F. (2012). Preventing depression: A global priority. *Journal of the American Medical Association*, *307*(10), 1033–1034. doi: 10.1001/jama.2012.271

Cuijpers, P., Van Straten, A., & Smit, F. (2005). Preventing the incidence of new cases of mental disorders: A meta-analytic review. *Journal of Nervous and Mental Disease*, *193*(2), 119–125. doi: 00005053-200502000-00006 [pii]

Deady, M., Choi, I., Calvo, R. A., Glozier, N., Christensen, H., & Harvey, S. B. (2017a). Ehealth interventions for the prevention of depression and anxiety in the general population: A systematic review and meta-analysis. *BMC Psychiatry*, *17*, 310. doi: 10.1186/s12888-017-1473-1

Deady, M., Johnston, D. A., Glozier, N., Milne, D., Choi, I., Mackinnon, A., … Harvey, S. B. (2018). Smartphone application for preventing depression: Study protocol for a workplace randomised controlled trial. *BMJ Open*, *8*(7), e020510. doi: 10.1136/bmjopen-2017-020510.

Deady, M., Peters, D., Lang, H., Calvo, R., Glozier, N., Christensen, H., & Harvey, S. B. (2017b). Designing smartphone mental health applications for emergency service workers. *Occupational Medicine*, *67*(6), 425–428. doi: 10.1093/occmed/kqx056

Fernandez, A., Salvador-Carulla, L., Choi, I., Calvo, R., Harvey, S. B., & Glozier, N. (2017). Development and validation of a prediction algorithm for the onset of common mental disorders in a working population. *Australian & New Zealand Journal of Psychiatry*, *52*(1), 47–58. doi: 10.1177/0004867417704506

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. doi: 10.1002/wps.20472

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. doi: 10.1111/2041-210X.12504

Harvey, S. B., Henderson, M., Lelliott, P., & Hotopf, M. (2009). Mental health and employment: Much work still to be done. *British Journal of Psychiatry*, *194*(3), 201–203. doi: 194/3/201 [pii]10.1192/bjp.bp.108.055111

Heber, E., Ebert, D. D., Lehr, D., Cuijpers, P., Berking, M., Nobis, S., & Riper, H. (2017). The benefit of web- and computer-based interventions for stress: A systematic review and meta-analysis. *Journal of Medical Internet Research*, *19*(2), e32. doi: 10.2196/jmir.5774

IBM Corp. (2015). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Kauer, S. D., Reid, S. C., Crooke, A. H. D., Khor, A., Hearps, S. J. C., Jorm, A. F., … 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. doi:10.2196/jmir.1858.

Kessler, R. C., Barber, C., Beck, A., Berglund, P., Cleary, P. D., McKenas, D., … Wang, P. (2003). The world health organization health and work performance questionnaire (HPQ). *Journal of Occupational and Environmental Medicine*, *45*(2), 156–174. doi: 10.1097/01.jom.0000052967.43131.51

Krieger, T., Zimmermann, J., Huffziger, S., Ubl, B., Diener, C., Kuehner, C., & Holtforth, M. G. (2014). Measuring depression with a well-being index: Further evidence for the validity of the WHO Well-Being Index (WHO-5) as a measure of the severity of depression. *Journal of Affective Disorders*, *156*, 240–244. doi: 10.1016/j.jad.2013.12.015

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9 – validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*, 317–325.

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. doi: 10.1002/wps.20673

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., … Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)*, *380*(9859), 2197–2223. doi: 10.1016/S0140-6736(12)61689-4

Peters, D., Deady, M., Glozier, N., Harvey, S., & Calvo, R. A. (2018). Worker preferences for a mental health app within male-dominated industries: Participatory study. *JMIR Mental Health*, *5*(2), e30. doi: 10.2196/mental.8999

Proudfoot, J., Parker, G., Hadzi Pavlovic, D., Manicavasagar, V., Adler, E., & Whitton, A. (2010). Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. *Journal of Medical Internet Research*, *12*(5), e64. doi: 10.2196/jmir.1475

Rathbone, A. L., Clarry, L., & Prescott, J. (2017). Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: Systematic review. *Journal of Medical Internet Research*, *19*(11), e399. doi: 10.2196/jmir.8598

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., … Finning, K. (2016). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. *The Lancet*, *388*(10047), 871–880. doi: 10.1016/S0140-6736(16)31140-0

Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, *32*(4), 329–342. doi: https://doi.org/10.1016/j.cpr.2012.02.004

Rose, G. (1994). *The strategy of preventive medicine*. New York: Oxford University Press.

Solomon, D., Proudfoot, J., Clarke, J., & Christensen, H. (2015). e-CBT (myCompass), antidepressant medication, and face-to-face psychological treatment for depression in Australia: A cost-effectiveness comparison. *Journal of Medical Internet Research*, *17*(11), e255. doi: 10.2196/jmir.4207

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *Journal of the American Medical Association*, *282*(18), 1737–1744.

Stratton, E., Choi, I., Lampit, A., Calvo, R. A., Harvey, S. B., & Glozier, N. (2017). Effectiveness of E-health interventions for reducing mental health

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press

conditions in employees: A systematic review and meta-analysis. *PLoS ONE*, *12*(12), e0189904. doi: https://doi.org/10.1371/journal.pone.0189904

Tan, L., Wang, M.-J., Modini, M., Joyce, S., Mykletun, A., Christensen, H., & Harvey, S. B. (2014). Preventing the development of depression at work: A systematic review and meta-analysis of universal interventions in the workplace. *BMC Medicine*, *12*, 74. doi: 10.1186/1741-7015-12-74

Tennant, C. (2001). Work-related stress and depressive disorders. *Journal of Psychosomatic Research*, *51*(5), 697–704. doi: https://doi.org/10.1016/S0022-3999(01)00255-0

Torous, J., & Roberts, L. (2017). Needed innovation in digital health and smartphone applications for mental health: Transparency and trust. *JAMA Psychiatry*, *74*(5), 437–438. doi: 10.1001/jamapsychiatry.2017.0262

Van Hooff, M., McFarlane, A. C., Davies, C. E., Searle, A. K., Fairweather-Schmidt, A. K., Verhagen, A., … Hodson, S. E. (2014). The Australian defence force mental health prevalence and wellbeing study: Design and methods. *European Journal of Psychotraumatology*, *5*, e23950. doi:10.3402/ejpt.v5.23950.

van Zoonen, K., Buntrock, C., Ebert, D. D., Smit, F., Reynolds, III, F. C., Beekman, A. T. F., & Cuijpers, P. (2014). Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *International Journal of Epidemiology*, *43*(2), 318–329. doi:10.1093/ije/dyt175.

Wang, P. S., Simon, G. E., Avorn, J., Azocar, F., Ludman, E. J., McCulloch, J., … Kessler, R. C. (2007). Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes: A randomized controlled trial. *Journal of the American Medical Association*, *298*(12), 1401–1411. doi: 10.1001/jama.298.12.1401

World Health Organization. (2017). Depression and other common mental disorders: global health estimates. Retrieved from http://www.webcitation.org/6ytV1RlIN.

https://doi.org/10.1017/S0033291720002081 Published online by Cambridge University Press