

## ORIGINAL ARTICLE

# The selfBACK artificial intelligence-based smartphone app can improve low back pain outcome even in patients with high levels of depression or stress

Guy Rughani<sup>1</sup>  | Tom I. L. Nilsen<sup>2</sup> | Karen Wood<sup>1</sup> | Frances S. Mair<sup>1</sup>  |  
 Jan Hartvigsen<sup>3,4</sup>  | Paul J. Mork<sup>2</sup>  | Barbara I. Nicholl<sup>1</sup> 

<sup>1</sup>School of Health and Wellbeing, University of Glasgow, Glasgow, UK

<sup>2</sup>Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

<sup>3</sup>Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

<sup>4</sup>Chiropractic Knowledge Hub, Odense, Denmark

## Correspondence

Barbara I. Nicholl, School of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, 1 Horselethill Road, Glasgow, G12 9LX, UK.

Email: [barbara.nicholl@glasgow.ac.uk](mailto:barbara.nicholl@glasgow.ac.uk)

## Abstract

**Background:** selfBACK provides individually tailored self-management support for low back pain (LBP) via an artificial intelligence-based smartphone app. We explore whether those with depressive/stress symptoms can benefit from this technology.

**Methods:** Secondary analysis of the selfBACK randomized controlled trial ( $n = 461$ ). Participants with LBP were randomized to usual care ( $n = 229$ ), or usual care plus selfBACK ( $n = 232$ ). Primary outcome: LBP-related disability (Roland–Morris Disability Questionnaire, RMDQ) over 9 months. Secondary outcomes: global perceived effect (GPE)/pain self-efficacy (PSEQ)/satisfaction/app engagement. Baseline depressive symptoms were measured using the patient health questionnaire (PHQ-8) and stress with the perceived stress scale (PSS). Outcomes stratified by baseline PHQ-8/PSS scores to assess associations across the whole cohort, and intervention versus control groups.

**Results:** Participants with higher levels of depressive/stress symptoms reported more baseline LBP-related disability (RMDQ 3.1; 1.6 points higher in most vs least depressed/stressed groups respectively); lower self-efficacy (PSEQ 8.1; 4.6 points lower in most vs least depressive/stressed groups respectively). LBP-related disability improved over time; relative risk of improvement in those with greatest depressive/stress symptoms versus nil symptom comparators at 9 months: 0.8 (95% CI: 0.6 to 1.0) and 0.8 (95% CI: 0.7 to 1.0) respectively. No evidence that different baseline levels of depressive/perceived stress symptoms are associated with different RMDQ/GPE/PSEQ outcomes. Whilst participants with higher PHQ-8/PSS were less likely to be satisfied or engage with the app, there was no consistent association among PHQ-8/PSS level, the intervention and outcomes.

**Conclusions:** The selfBACK app can improve outcomes even in those with high levels of depressive/stress symptoms and could be recommended for patients with LBP.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *European Journal of Pain* published by John Wiley & Sons Ltd on behalf of European Pain Federation - EFIC®.

**Significance:** We have demonstrated that an app supporting the self-management of LBP is helpful, even in those with higher levels of baseline depression and stress symptoms. selfBACK offers an opportunity to support people with LBP and provides clinicians with an additional tool for their patients, even those with depression or high levels of stress. This highlights the potential for digital health interventions for chronic pain.

## 1 | INTRODUCTION

Low back pain (LBP) is the greatest cause of global disability (Hoy et al., 2014), and is the condition most commonly associated with other chronic conditions (Schäfer et al., 2014). Psychological factors are frequently associated with recurrent episodes of LBP, but the mechanisms are complex (Hartvigsen et al., 2018; Pinheiro et al., 2016; Stubbs et al., 2016). Neurobiological underpinnings, including neurocognitive processing, have been shown to play a role in the connection between chronic pain and poor mood and distress (Rusu et al., 2019), contributing to the complexity of chronic pain and the unique experience lived by each individual (Clauw et al., 2019).

A systematic review, including 13 longitudinal studies, found that depression was associated with increased disability in LBP (Pinheiro et al., 2016). Depression also appears to increase the incidence of future chronic LBP (Currie & Wang, 2005).

Perceived stress has also been associated with LBP development and duration (Hartvigsen et al., 2018; Power et al., 2001). A review of 16 observational studies reported that catastrophizing was associated with delayed recovery from LBP (Wertli et al., 2014). The aligned, but distinct concept of feeling overwhelmed by long-term nonspecific stressors has also been associated with LBP intensity and pain-related disability (Puschmann et al., 2020).

Self-management strategies are recommended in LBP guidelines, where the patient is encouraged to learn about and manage their condition (Bernstein et al., 2017; Foster et al., 2018). Digital interventions (websites, mobile applications (apps), wearable technology) are seen as engaging and scalable ways to deliver this information (Dennison et al., 2013; Zhao et al., 2016). However, self-management requires motivation and confidence – attributes that may be difficult for those with depressive symptoms or high levels of perceived stress.

The selfBACK digital intervention is an artificial intelligence (AI)-based mobile app designed to generate evidence-based, individually tailored self-management support of non-specific LBP (Mork & Bach, 2018). The selfBACK randomized controlled trial (RCT) found that as an adjunct to usual care, selfBACK resulted in lower

LBP-related disability at 3 months compared to usual care alone, and its benefits were sustained throughout the 9-month follow-up period (Sandal et al., 2021).

In this secondary analysis, we aim to investigate whether depressive symptoms and perceived stress influence outcomes for people with LBP in the selfBACK RCT. The specific research questions were as follows:

- Q1 Is there a difference at baseline in LBP-related disability and self-efficacy between those with high versus low levels of depressive or perceived stress symptoms?
- Q2 What is the trajectory of LBP-related disability by baseline depressive symptom score or level of perceived stress over a 9-month period?
- Q3 Is the effect of the selfBACK app on LBP-related disability, global perceived effect and self-efficacy at 3 and 9 months modified by the level of baseline depressive symptoms or level of perceived stress?
- Q4 Are people with high baseline depressive symptoms or perceived stress less likely to be satisfied with the selfBACK app than those without?
- Q5 Are people with high baseline depressive symptoms or perceived stress less likely to engage with the selfBACK app than those without?

## 2 | METHODS

### 2.1 | Study design, setting and participants

This was a secondary analysis of the randomized, assessor-blinded international multicentre trial of the selfBACK app for patients with non-specific LBP (Trial Registration: NCT03798288. Date of registration: 9 January 2019; <https://clinicaltrials.gov/ct2/show/NCT03798288>). We explored the influence of depressive symptoms and perceived stress at baseline on selected outcomes of the main RCT. The full protocol and results of the RCT are published elsewhere (Mork & Bach, 2018; Sandal et al., 2019, 2021). In summary, patients who had attended their primary care provider or an outpatient spine clinic with LBP

within the previous 8 weeks were invited to complete a web-based questionnaire. A total of 461 patients from primary care (Denmark and Norway) and an outpatient spine clinic (Denmark) were randomized to usual care ( $n = 229$ ) or usual care plus selfBACK app ( $n = 232$ ). LBP-related disability, assessed by the Roland–Morris Disability Questionnaire (RMDQ) (Roland & Fairbank, 2000), was the primary outcome; a score of 6 or above was required to be eligible for the trial. Potential participants were 18 years or older, required email, computer and smartphone access, could speak, read or understand Danish or Norwegian and had no cognitive impairment, learning disability or conditions limiting participation, contraindication to exercise or physical activity, fibromyalgia, pregnancy, previous back surgery or ongoing participation in other LBP management trials.

Those randomized to the selfBACK intervention received weekly, individually tailored, recommendations for physical activity, strength and flexibility exercises and daily educational messages via the app. User data from symptom progression, step count, exercise completion and questionnaire information fed back to the case-based management system to tailor recommendations based on what had successfully worked in cases with similar characteristics and symptoms (Bach et al., 2016). Usual care meant managing LBP as per the advice or treatment from their care provider.

## 2.2 | Depressive and stress symptom measures

Sociodemographic information, depressive symptoms and perceived stress were collected at baseline. Baseline depressive symptoms were measured by the Patient Health Questionnaire (PHQ-8). The PHQ-8 is a widely accepted and validated instrument for assessing depressive symptoms in epidemiological studies (Arias-de la Torre et al., 2021; Wu et al., 2020). It asks about eight items related to the DSM-IV diagnostic criteria for depression over the previous 2-week period and responses are on a scale from 0 (not at all) to 3 (nearly every day) which are summed to give a total PHQ-8 score of 0–24; higher scores indicate greater depressive symptoms (Kroenke et al., 2009).

Perceived stress was measured with the 10-item Perceived Stress Scale (PSS), in which participants rated how upsetting, uncontrollable, unpredictable and overwhelming their lives had been during the previous month (Cohen et al., 1983; Cohen & Williamson, 1988; Lee, 2012). The scale was developed for use in community-based samples and not as a diagnostic instrument (Cohen et al., 1983), it has been shown to have good reliability

and to correlate with other measures of stress (Cohen & Williamson, 1988). Responses were rated on a 0 (never) to 4 (very often) point scale; with four positively stated items reverse scored, the 10 responses were then summed to give a total PSS score of 0–40, with higher scores indicating greater perceived stress (Cohen & Williamson, 1988).

## 2.3 | Outcomes

RMDQ was measured at baseline, 6 weeks, 3, 6 and 9 months using a web-based questionnaire. Higher scores (0–24) indicated higher LBP-related disability. The mean difference in RMDQ between the intervention and control group at 3 months was the primary outcome for the RCT. Although there is some debate in the literature, a clinically meaningful improvement in RMDQ score was regarded as a decrease of  $\geq 4$  points (Jordan et al., 2006; Ostelo & de Vet, 2005).

Secondary outcomes were (1) the 10-item Pain Self-Efficacy Questionnaire (PSEQ: range 0–60), which has shown strong psychometric properties in LBP populations, participants are asked to rate their confidence in taking part in activities despite their pain (Nicholas, 2007). Higher scores indicate greater confidence in the ability to cope despite pain; and (2) the Global Perceived Effect (GPE: range –5 to 5) scale estimate of overall improvement, where positive scores indicate LBP improvement, and negative scores a deterioration (Kamper et al., 2009). PSEQ was recorded at baseline; both PSEQ and GPE were recorded at 6 weeks, 3, 6 and 9 months.

Satisfaction with the app was measured at a single point at 4 months on a 5-point Likert scale (1 lowest and 5 highest satisfaction) by asking the participants: ‘how do you rate the selfBACK app?’ For analysis, ‘satisfied’ participants were those scoring 4 or 5.

App engagement was measured by the number of weekly self-management plans participants created on the app. ‘Engaged’ participants were defined as those creating 6 or more self-management plans during the first 3 months post-randomization.

## 2.4 | Statistical analysis

PHQ-8 respondents were categorized into three groups that balanced the score's recognized clinical categories (Kroenke et al., 2001) and the number of trial participants: no depressive symptoms (0–4 score,  $n = 180$ ), mild depressive symptoms (5–9 score,  $n = 188$ ) and moderate, moderately severe and severe depressive symptoms (10–24 score,  $n = 93$ ). For analyses of the modifying effect, participants were classified as having nil and mild (0–9

score,  $n = 368$ ) vs moderate, moderately severe and severe depressive symptoms (10–24 score,  $n = 93$ ). A PHQ-8 score  $\geq 10$  can also be interpreted as ‘current depression’ (Kroenke et al., 2009).

PSS respondents were categorized for all analyses into two groups based on the score’s clinical categories (Cohen et al., 1983) and the number of trial participants: nil and low stress (0–13 score,  $n = 191$ ) and moderate and high stress (14–40 score,  $n = 270$ ).

To address Q1, we described the cohort’s baseline RMDQ and PSEQ scores stratified by baseline PHQ-8 or PSS level. The crude and adjusted differences from the no depression or nil/low stress reference categories are presented.

For Q2, we presented the trajectory of RMDQ at baseline, 3 and 9 months and calculated the crude and adjusted mean change in RMDQ from baseline stratified by PHQ-8 or PSS category using a linear mixed model. We also repeated this analysis to consider PHQ-8 and PSS scores as a continuous variable. Further, we calculated the relative risk for an improvement of  $\geq 4$  points in RMDQ at 3 and 9 months, compared to the no depressive symptoms or nil/low perceived stress groups using a Poisson generalized estimated equation (GEE) model.

For Q3, we estimated the effect of the intervention using a constrained longitudinal data analyses to approach as described in the primary outcome paper (Sandal et al., 2021), and presented RMDQ, GPE and PSEQ stratified by PHQ-8 or PSS category. This implies using a common baseline category for the intervention and control group, and model group differences at each follow-up time point in a linear mixed model. Measures of effect modification by depression or stress were then obtained from a post-estimation command that calculated the difference between the strata-specific effects at 3 and 9 months with associated  $p$ -values. Throughout, statistical significance was defined as  $p$ -value  $< 0.05$ .

Q4 analysed data from the intervention group using logistic regression to estimate the adjusted odds ratio with 95% CI of being ‘satisfied’ with the app (defined above) in participants with depression or stress, compared to reference groups with no depressive symptoms or nil/low perceived stress.

Q5 was analysed as per Q4 but calculated the adjusted odds ratio with 95% CI of being ‘engaged’ with the app (defined above).

The main trial was powered at 90% to detect a difference in RMDQ of  $\geq 2$  points at 3 months’ follow-up with a planned sample size of 350 participants (Sandal et al., 2021). As in the main trial, analyses were adjusted for country of recruitment, recruiting clinician, education ( $< 10$ , 10–12,  $> 12$  years), duration of current pain episode ( $< 1$ , 1–4, 5–12,  $> 12$  weeks), average pain intensity

in the preceding week (continuous, range 0–10 scale), sex (male, female) and age (years). When analysing the trajectory of the whole cohort (research Q2), we also adjusted for baseline body mass index, work-ability (11-point self-rated scale from 0: unable to work to 10: fully able to work [Ahlstrom et al., 2010]) and physical activity level (self-reported time per week performing leisure activities with a revised version of the Saltin–Grimby score: level 1: sedentary to level 4 regular vigorous activity [Grimby et al., 2015]).

All analyses were performed using Stata version 16.1.

## 2.5 | Ethics

All participants provided informed written consent before commencing the RCT, which was approved by the Danish Data Protection Agency (201–57-0008) and regional ethics committees in Denmark (S-20182000-24) and Norway (2017/923–6).

## 3 | RESULTS

Table 1 summarizes the characteristics of the study population, and baseline PHQ-8, PSS and PSEQ scores. Of the 461 participants enrolled, complete RMDQ data were obtained for 368 (80%) at 6 weeks, 399 (87%) at 3 months, 349 (76%) at 6 months and 352 (76%) at 9 months (Sandal et al., 2021).

### 3.1 | Research Q1

Analysing the intervention and control groups as a whole cohort, when stratified by baseline PHQ-8 or PSS categories, participants with higher depressive symptoms or higher perceived stress scores reported higher adjusted baseline RMDQ scores (Table 2). Groups with higher levels of depressive or stress symptoms also had less confidence in their ability to cope with pain: adjusted difference in PSEQ between moderate to severe and no depressive symptoms:  $-8.1$  (95% CI:  $-10.7$  to  $-5.4$ ), and between moderate to high stress and nil to low stress symptoms:  $-4.6$  (95% CI:  $-6.5$  to  $-2.7$ ).

Extending this analysis to consider PHQ-8 and PSS as continuous variables (Table 2) showed that for each unit increase in PHQ-8 and PSS scores, there was an increase of 0.3 (95% CI: 0.2–0.4) and 0.1 (95% CI: 0.1–0.2) of the RMDQ score respectively; and a reduction in the PSEQ score (i.e., poorer pain self-efficacy) of  $-0.8$  (95% CI:  $-1.0$  to  $-0.5$ ) and  $-0.4$  (95% CI:  $-0.6$  to  $-0.3$ ) respectively.

**TABLE 1** Baseline characteristics of the study population.

|  | All (n = 461)       | Usual care (n = 229)    | selfBACK (n = 232)  |
|--|---------------------|-------------------------|---------------------|
| Age in years, mean (SD), [range]             | 47.5 (14.7) [18–86] | 46.7 (14.4) [18–81]     | 48.3 (15.0) [20–86] |
| Sex, women, n (%)                            | 255 (55.3)          | 134 (58.5)              | 121 (52.2)          |
| BMI kg/m <sup>2</sup> , mean (SD), [range]   | 27.6 (5.1) [17–54]  | 27.8 (5.4) [18–54]      | 27.3 (4.7) [17–46]  |
| Full-time employment, n (%)                  | 281 (61.0)          | 143 (62.4)              | 138 (59.5)          |
| Married/living with partner, n (%)           | 332 (72.0)          | 158 (69.0)              | 174 (75.0)          |
| Clinical setting of patient recruitment      |                     |                         |                     |
| Primary care, n (%)                          | 68 (14.8)           | 34 (14.8)               | 34 (14.7)           |
| Physiotherapist, n (%)                       | 135 (29.3)          | 67 (29.3)               | 68 (29.3)           |
| Chiropractor, n (%)                          | 160 (34.7)          | 79 (34.5)               | 81 (34.9)           |
| Outpatient clinic, n (%)                     | 98 (21.3)           | 49 (21.4)               | 49 (21.1)           |
| Education, >12 years, n (%)                  | 297 (64.4)          | 145 (63.3) <sup>f</sup> | 152 (65.5)          |
| SGPALS                                       |                     |                         |                     |
| Sedentary, n (%)                             | 33 (7.2)            | 18 (7.9)                | 15 (6.5)            |
| Some physical activity, n (%)                | 239 (51.8)          | 121 (52.8)              | 118 (50.9)          |
| Pain duration, ≥12 weeks, n (%)              | 267 (57.9)          | 136 (59.4)              | 131 (56.5)          |
| Depressive symptoms: PHQ-8 (categorized)     |                     |                         |                     |
| Nil 0–4, n (%)                               | 180 (39.1)          | 77 (33.6)               | 103 (44.4)          |
| Mild 5–9, n (%)                              | 188 (40.8)          | 104 (45.4)              | 84 (36.2)           |
| Moderate 10–14, n (%)                        | 68 (14.8)           | 37 (16.2)               | 31 (13.4)           |
| Moderately severe 15–19, n (%)               | 20 (4.3)            | 9 (3.9)                 | 11 (4.7)            |
| Severe 20–24, n (%)                          | 5 (1.1)             | 2 (0.9)                 | 3 (1.3)             |
| Perceived stress symptoms: PSS (categorized) |                     |                         |                     |
| Low 0–13, n (%)                              | 191 (41.4)          | 89 (38.9)               | 102 (44.0)          |
| Moderate 14–26, n (%)                        | 249 (54.0)          | 130 (56.8)              | 119 (51.3)          |
| High 27–40, n (%)                            | 21 (4.6)            | 10 (4.4)                | 11 (4.6)            |
| RMDQ, mean (SD)                              | 10.4 (4.4)          | 10.6 (4.4)              | 10.3 (4.4)          |
| PSEQ, mean (SD)                              | 44.1 (11.1)         | 45.3 (10.4)             | 42.8 (11.2)         |
| GPE  | N/A                 | N/A                     | N/A                 |

Abbreviations: BMI, body mass index, GPE, global perceived effect, PHQ-8, patient health questionnaire 8, PSEQ, pain self-efficacy questionnaire, PSS, perceived stress scale, RMDQ, Roland–Morris disability questionnaire, SPGALS, Saltin–Grimby physical activity level scale.

### 3.2 | Research question 2

Across the whole cohort, there was an improvement in RMDQ score for those with all levels of baseline depressive symptoms and perceived stress scores at 9 months' follow-up, although all confidence intervals were wide and crossed zero (Tables 3 and 4; Tables S1 and S2).

Level of baseline depression or perceived stress was inversely associated with the probability of a clinically significant improvement in LBP-related disability (RMDQ change ≥4 points) at 3 months. Compared to people with no depressive symptoms, those who reported mild, or moderate-to-severe symptoms had a RR of 0.8 (95% CI: 0.6–0.9) and 0.7 (95% CI: 0.5–1.0) respectively. Compared to those with nil or low stress, those reporting moderate or high stress had a RR of

0.8 (95% CI: 0.7–1.0). These trends were maintained for 9 months.

### 3.3 | Research question 3

Regardless of the level of depressive symptoms or perceived stress, those receiving the selfBACK intervention had better outcomes in RMDQ, GPE and PSEQ than those receiving usual care alone at 3 and 9 months; however, this was not statistically significant (Table 5). The exception was the RMDQ score at 9 months in those with moderate–severe depression randomized to selfBACK, which was similar to the RMDQ score of those with moderate–severe depression randomized to usual care at this timepoint (8.8 vs. 8.7, respectively).

**TABLE 2** LBP-related disability (RMDQ) and self-efficacy (PSEQ) at baseline for all participants stratified by baseline depression symptoms (PHQ) and perceived stress level (PSS).

| Variable                                 | RMDQ       |                  |   | PSEQ        |                  |   |
|--|------------|------------------|---|-------------|------------------|---|
|  | Mean (SD)  | Crude difference | Adjusted <sup>a</sup> difference (95% CI) | Mean (SD)   | Crude difference | Adjusted <sup>a</sup> difference (95% CI) |
| Depressive symptoms                      |            |                  |   |             |                  |   |
| No (PHQ 0–4)                             | 8.9 (3.8)  | 0.0              | 0.0 (Reference)                           | 47.7 (9.3)  | 0.0              | 0.0 (Reference)                           |
| Mild (PHQ 5–9)                           | 10.6 (4.2) | 1.7              | 1.6 (0.8 to –2.3)                         | 43.8 (10.9) | –3.9             | –3.8 (–5.8 to –1.8)                       |
| Moderate or severe (PHQ 10–24)           | 13.0 (4.7) | 4.1              | 3.1 (2.1 to 4.1)                          | 38.0 (11.7) | –9.7             | –8.1 (–10.7 to –5.4)                      |
| Depressive symptoms, per unit (PHQ 0–24) | 10.4 (4.4) | 0.4              | 0.3 (0.2 to 0.4)                          | 44.1 (11.0) | –0.9             | –0.8 (–1.0 to –0.5)                       |
| Perceived stress                         |            |                  |   |             |                  |   |
| Nil or low (PSS 0–13)                    | 9.1 (4.0)  | 0.0              | 0.0 (Reference)                           | 47.4 (10.0) | 0.0              | 0.0 (Reference)                           |
| Moderate or high (PSS 14–40)             | 11.4 (4.4) | 2.2              | 1.6 (0.9 to 2.3)                          | 41.8 (11.1) | –5.6             | –4.6 (–6.5 to –2.7)                       |
| Perceived stress, per unit (PSS 0–40)    | 10.4 (4.4) | 0.2              | 0.1 (0.1 to 0.2)                          | 44.1 (11.0) | –0.5             | –0.4 (–0.6 to –0.3)                       |

Abbreviations: PHQ-8, patient health questionnaire 8; PSEQ, pain self-efficacy questionnaire; PSS, perceived stress scale; RMDQ, Roland–Morris disability questionnaire.

<sup>a</sup>Adjusted for country, recruiting clinician, education, pain duration at baseline, pain intensity at baseline, sex, age, body mass index, work ability score at baseline and physical activity at baseline.

The overall trend suggested that those with more severe depressive or perceived stress symptoms in the selfBACK groups had marginally greater improved RMDQ, GPE and PSEQ scores at 3 months compared to those in the control groups or those with less severe symptoms. At 9 months, this trend was maintained for GPE in the depressive symptom groups, but reversed in all other outcomes, such that those with less severe symptoms fared better.

### 3.4 | Research question 4

Analysing the intervention group only, participants with mild- or moderate-to-severe depressive symptoms were less likely to be satisfied with the app compared to those with no depressive symptoms (adjusted OR 0.5, 95% CI: 0.4–0.7 and OR 0.7, 95% CI: 0.4–1.1 respectively). The same was found for perceived stress, where those with moderate and high stress were less likely to be satisfied with the app compared to those with nil and low perceived stress (adjusted OR 0.5, 95% CI: 0.4–0.7).

### 3.5 | Research question 5

Analysing the intervention group only, those with mild- or moderate-to-severe depressive symptoms were less likely

to engage with the app compared to a reference group with no depressive symptoms (both groups adjusted OR 0.6, 95% CI: 0.5–0.8). Similarly, those with moderate and high perceived stress were less likely to engage with the app compared to those with nil and low perceived stress (adjusted OR 0.8, 95% CI: 0.6–0.9).

## 4 | DISCUSSION AND CONCLUSIONS

### 4.1 | Summary of findings

People with higher levels of depressive symptoms or perceived stress reported higher levels of LBP-related disability (RMDQ) and lower pain self-efficacy (PSEQ) at baseline. Over 9 months' follow-up, during which the cohort either received usual care or usual care plus selfBACK, LBP-related disability tended to improve. However, the probability of improvement decreased with increasing levels of baseline depressive symptoms or perceived stress.

Those in the intervention group with any level of elevated baseline depressive or stress symptoms tended towards better outcomes in RMDQ, GPE and PSEQ than those receiving usual care alone. There was no evidence that different baseline levels of depressive or perceived stress symptoms were associated with different RMDQ,

**TABLE 3** The trajectory of LBP-related disability (RMDQ) over 9-month follow-up stratified by level of depressive symptoms at baseline (PHQ-8).

| Depressive symptom level                        | Mean RMDQ (SD)                   |            |           | Change in RMDQ from baseline to 3 months |   | Change in RMDQ from baseline to 9 months |   |
|---|----------------------------------|------------|-----------|--|---|--|---|
|   | Baseline                         | 3 months   | 9 months  | Crude mean                               | Adjusted <sup>a</sup> difference (95% CI) | Crude mean                               | Adjusted <sup>a</sup> difference (95% CI) |
|   | None (PHQ-8: 0–4) <i>n</i> = 180 | 9.2 (3.8)  | 5.8 (4.2) | 5.4 (4.5)                                | −3.4                                      | 0.0 (Reference)                          | −3.8                                      |
| Mild (PHQ-8: 5–9) <i>n</i> = 188                | 10.9 (4.2)                       | 7.4 (5.1)  | 6.9 (5.7) | −3.4                                     | −0.1 (−0.9 to 0.9)                        | −3.9                                     | −0.1 (−1.0 to 0.9)                        |
| Moderate to severe (PHQ-8: 10–24) <i>n</i> = 93 | 13.3 (4.7)                       | 10.0 (5.4) | 9.0 (6.2) | −3.3                                     | 0.1 (−1.0 to 1.3)                         | −4.3                                     | −0.5 (−1.7 to 0.7)                        |

Abbreviations: PHQ-8, patient health questionnaire 8; RMDQ, Roland–Morris disability questionnaire.

<sup>a</sup>Adjusted for country, recruiting clinician, education, pain duration at baseline, pain intensity at baseline, sex, age, body mass index, work ability score at baseline and physical activity at baseline.

GPE or PSEQ outcomes. Those with moderate to severe depressive symptoms or high perceived stress at baseline were less likely to be satisfied or engage with the app than those with no depressive symptoms or nil/low perceived stress.

## 4.2 | Comparison with existing literature

To our knowledge, this is the first study to explore how depressive symptoms or perceived stress may influence the acceptability and effectiveness of a digital intervention to promote self-management of LBP. In line with existing literature, participants with higher levels of depressive symptoms and perceived stress report higher levels of LBP-related disability and lower self-efficacy (Bair et al., 2009; Jackson et al., 2014; Puschmann et al., 2020).

Different trajectories of non-specific LBP have been noted and attempts have been made to categorize patients into subgroups based on clinical characteristics (Axén & Leboeuf-Yde, 2013). For example, there is a growing literature suggesting an association of depression with greater chronicity (Bair et al., 2009; Currie & Wang, 2005) and sustained higher levels of pain-related disability and functional disability (Andersen et al., 2022), although cause and effect may be debated. Previous research has suggested that psychologically augmented physiotherapy is effective for those with high levels of pain catastrophizing or fear avoidance (Hill et al., 2011). Unfortunately, post-COVID-19 pandemic access to rapid and appropriate physiotherapy is suboptimal in many countries (Equipsme, 2022). Digital health interventions have the potential to promote more rapid access to appropriate self-management support for LBP. Importantly for primary care, where depression, stress and non-specific LBP are highly co-prevalent (Pincus & McCracken, 2013), we have demonstrated that those with severe depressive or stress symptoms can benefit from digital self-management support for LBP.

Our work has also shown that non-specific LBP tends to improve over the medium term (9 months), resonating with work by Axén et al. (2011) that looked at the 6-month course of LBP but differs somewhat from a previous study by Dunn et al. (2006), which suggested that those with poor psychological status were more likely to experience severe, chronic pain and disability over the course of 12 months. However, these results should be interpreted with caution since psychological symptoms in people with pain may appear qualitatively different from symptoms in other groups (Rusu & Hallner, 2018; Rusu & Pincus, 2017).

As hypothesized, engagement and satisfaction were lower in those with greater levels of depression or stress. Maintaining engagement with app-delivered interventions is particularly challenging in those with underlying mental

**TABLE 4** The trajectory of LBP-related disability (RMDQ) over 9-month follow-up stratified by level of perceived stress symptoms at baseline (PSS).

| Perceived stress level                               | Mean RMDQ (SD) |           |           | Change in RMDQ from baseline to 3 months |   | Change in RMDQ from baseline to 9 months |   |
|--|----------------|-----------|-----------|--|---|--|---|
|  | Baseline       | 3 months  | 9 months  | Crude mean                               | Adjusted <sup>a</sup> difference (95% CI) | Crude mean                               | Adjusted <sup>a</sup> difference (95% CI) |
| Nil and low stress (PSS: 0–13) <i>n</i> = 191        | 9.5 (4.0)      | 6.3 (4.8) | 5.8 (5.1) | –3.2                                     | 0.0 (Reference)                           | –3.7                                     | 0.0 (Reference)                           |
| Moderate and high stress (PSS: 14–40) <i>n</i> = 270 | 11.7 (4.4)     | 8.1 (5.2) | 7.5 (5.7) | –3.5                                     | –0.3 (–1.2 to 0.5)                        | –4.2                                     | –0.5 (–1.3 to 0.4)                        |

Abbreviations: PSS, perceived stress scale, RMDQ: Roland–Morris disability questionnaire.

<sup>a</sup>Adjusted for country, recruiting clinician, education, pain duration at baseline, pain intensity at baseline, sex, age, body mass index, work ability score at baseline and physical activity at baseline.

health problems (Linardon & Fuller-Tyszkiewicz, 2020) but selfBACK's ability to personalize content to each user meets best-practice recommendations (Schubart et al., 2011), suggesting that this may be a useful additional tool for primary care practitioners managing those with LBP.

### 4.3 | Strengths and limitations

The main trial was robustly conducted, with effective randomization, a person-centred intervention and multi-centred, international recruitment from a variety of clinical settings (Sandal et al., 2021). Despite the stated risks of loss to follow-up, especially in the groups with higher levels of depressive symptoms or perceived stress, there was little attrition (Sandal et al., 2021). This may reflect the usability and perceived utility of the intervention.

However, this was a secondary analysis of a trial not originally powered to detect differences by depressive symptom score or levels of perceived stress. PHQ-8 and PSS were both measured on approximately continuous scales. The variables were categorized to allow for possible non-linear associations and to address effect modification in stratified analyses. The number of categories and their cut-offs were selected as a necessary compromise between clinical relevance and meaningful statistical power. The small numbers limit the precision of the estimates as reflected by the wide confidence intervals throughout, suggesting that results must be interpreted with caution. To increase statistical power, we also analysed PHQ-8 and PSS scores as continuous variables and observed weak positive associations with RMDQ and weak inverse associations with PSEQ score. Although this suggests a dose–response relationship of PHQ-8 and PSS with both RMDQ and PSEQ, these are exploratory analyses and do not necessarily infer causality. The causal mechanisms underpinning the complex relationships between pain and mood require further elucidation. It may be that people with high levels of depressive symptoms or perceived stress need further support to help with their mental health before tackling self-management of LBP using selfBACK.

There is some debate about the best measure for LBP-related disability (Chiarotto et al., 2016; Kent et al., 2015). The RMDQ has been shown to have adequate measurement properties in a range of domains as well as good test–retest reliability (Jenks et al., 2022; Lauridsen et al., 2006); it is particularly suited to studies in primary care populations (Lauridsen et al., 2006) and is one of the recommended core outcome measures for use in non-specific LBP RCTs (Chiarotto et al., 2018). Measurement error is at random (Jenks et al., 2022) and therefore any bias is likely to result in an underestimate of pain-related disability, so we are confident that the relationships reported in this study are not exaggerated.



**TABLE 5** The effect of selfBACK on LBP-related disability (RMDQ), perceived effect (GPE) and self-efficacy (PSEQ) at 3 and 9 months by the level of baseline depressive (PHQ-8) and perceived stress symptoms (PSS).

|          | Depressive symptom level                        |  |   |  | Perceived stress level                       |  |   |  |
|----------|---|--|---|--|--|--|---|--|
|          | Nil and mild depression (PHQ-8: 0–9)<br>n = 368 |  | Moderate, moderately severe and severe depression (PHQ-8: 10–24) n = 93 |  | Nil and low stress (PSS: 0–13) n = 191       |  | Moderate and high stress (PSS: 14–40) n = 270 |  |
|          | Crude mean (SD)                                 | Adjusted <sup>a</sup> mean difference (95% CI) | Crude mean (SD)   | Adjusted <sup>a</sup> mean difference (95% CI) | Crude mean (SD)                              | Adjusted <sup>a</sup> mean difference (95% CI) | Crude mean (SD)                               | Adjusted <sup>a</sup> mean difference (95% CI) |
|          | Difference in adjusted means (P interaction)    |  |   |  | Difference in adjusted means (P interaction) |  |   |  |
|          | Control (n = 181)                               | selfBACK (n = 187)                             | Control (n = 48)  | selfBACK (n = 45)                              | Control (n = 89)                             | selfBACK (n = 102)                             | Control (n = 140)                             | selfBACK (n = 130)                             |
| RMDQ     |   |  |   |  |  |  |   |  |
| Baseline | 9.8 (4.1)                                       |  | 13.0 (4.7)  |  | 9.1 (4.0)                                    |  | 11.4 (4.4)                                    |  |
| 3 months | 6.6 (4.9)                                       | -0.6 (-1.4 to 0.2)                             | 10.7 (6.1)  | 8.9 (4.5)                                      | 6.1 (4.9)                                    | 5.8 (4.7)                                      | 8.3 (5.6)                                     | 7.3 (4.7)                                      |
| 9 months | 6.4 (5.4)                                       | -1.0 (-1.84 to -0.20)                          | 8.7 (6.4)   | 8.8 (6.1)                                      | 6.2 (5.5)                                    | 4.8 (4.5)                                      | 7.3 (5.7)                                     | 7.0 (5.6)                                      |
| GPE      |   |  |   |  |  |  |   |  |
| Baseline | N/A   | N/A  | N/A   | N/A  | N/A  | N/A  | N/A   | N/A  |
| 3 months | 1.3 (1.9)                                       | 0.6 (0.2 to 1.0)                               | 1.0 (1.9)   | 1.9 (1.9)                                      | 1.4 (1.8)                                    | 2.0 (2.0)                                      | 1.2 (2.0)                                     | 1.9 (1.9)                                      |
| 9 months | 1.4 (2.2)                                       | 0.7 (0.3 to 1.2)                               | 1.1 (2.2)   | 2.0 (1.8)                                      | 1.4 (2.2)                                    | 2.3 (2.1)                                      | 1.3 (2.2)                                     | 2.0 (1.9)                                      |
| PSEQ     |   |  |   |  |  |  |   |  |
| Baseline | 45.7 (10.3)                                     |  | 38.0 (11.7)   |  | 47.4 (10.0)                                  |  | 41.8 (11.2)                                   |  |
| 3 months | 48.2 (10.6)                                     | 2.3 (0.8 to 3.9)                               | 40.5 (11.8)   | 43.6 (11.3)                                    | 49.4 (9.3)                                   | 51.9 (9.1)                                     | 44.6 (11.9)                                   | 47.2 (10.3)                                    |
| 9 months | 48.0 (10.4)                                     | 3.7 (2.0 to 5.3)                               | 42.8 (12.9)   | 44.0 (12.1)                                    | 48.9 (10.5)                                  | 53.6 (7.2)                                     | 45.5 (11.2)                                   | 47.6 (10.9)                                    |

Abbreviations: GPE, global perceived effect, PHQ-8, patient health questionnaire 8, PSEQ, pain self-efficacy questionnaire, PSS: perceived stress scale, RMDQ, Roland-Morris disability questionnaire.

<sup>a</sup>Adjusted for country, recruiting clinician, education, pain duration at baseline, pain intensity at baseline, sex and age.

Although there were participants with severe depressive symptom scores and high perceived stress scores, it is possible that the very fact they participated in the trial may mean that they represent a more motivated cohort than the general LBP population with depression and stress (Axén & Leboeuf-Yde, 2013). The requirement for access to the internet and a smartphone also potentially limits the findings to those of higher socioeconomic status.

#### 4.4 | Implications for practice, policy, education and future research

Given the high prevalence of depressive and stress symptoms, future digital interventions for LBP should record these factors at baseline and follow up and explore the effects, if any, of antidepressant/anxiolytic use. Self-efficacy should also be monitored, to better establish its interplay with LBP and depression/stress (Puschmann et al., 2020). For primary care clinicians, this work suggests that even patients with depressive symptoms or high levels of stress should be considered for digital interventions for LBP as they are still likely to benefit. The study highlights the potential for AI-based apps to tailor self-management support for chronic illness, which may be particularly useful for people with combined physical and mental health problems.

## 5 | CONCLUSION

App-based interventions may improve outcomes even in those with high levels of depressive symptoms and perceived stress and could be recommended for this population of patients with LBP.

### AUTHOR CONTRIBUTIONS

BIN, FSM, KW, JH, PJM and TILN were involved in the conception of this study and contributed to the study design. Statistical analyses were carried out by TILN. GR drafted the manuscript. All authors interpreted the results and commented on the manuscript before approving the final version.

### ACKNOWLEDGEMENTS

The authors would like to thank selfBACK participants for taking part in the study and the full selfBACK study team of investigators and researchers.

### FUNDING INFORMATION

This work was supported by European Union Horizon 2020 research and innovation program under grant agreement number 689043.

### CONFLICT OF INTEREST

None declared.

### ORCID

Guy Rughani  <https://orcid.org/0000-0002-3359-244X>  
 Frances S. Mair  <https://orcid.org/0000-0001-9780-1135>  
 Jan Hartvigsen  <https://orcid.org/0000-0002-5876-7410>  
 Paul J. Mork  <https://orcid.org/0000-0003-3355-2680>  
 Barbara I. Nicholl  <https://orcid.org/0000-0001-5639-0130>

### REFERENCES

- Ahlstrom, L., Grimby-Ekman, A., Hagberg, M., & Dellve, L. (2010). The work ability index and single-item question: Associations with sick leave, symptoms, and health—a prospective study of women on long-term sick leave. *Scandinavian Journal of Work, Environment & Health*, 36, 404–412.
- Andersen, T. E., Karstoft, K. I., Lauridsen, H. H., & Manniche, C. (2022). Trajectories of disability in low back pain. *Pain Reports*, 7, e985.
- Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Serrano-Blanco, A., & Alonso, J. (2021). PHQ-8 scores and estimation of depression prevalence - Author's reply. *Lancet Public Health*, 6, e794.
- Axén, I., Bodin, L., Bergström, G., Halasz, L., Lange, F., Lövgren, P. W., Rosenbaum, A., Leboeuf-Yde, C., & Jensen, I. (2011). Clustering patients on the basis of their individual course of low back pain over a six month period. *BMC Musculoskeletal Disorders*, 12, 1–10.
- Axén, I., & Leboeuf-Yde, C. (2013). Trajectories of low back pain. *Best Practice & Research. Clinical Rheumatology*, 27, 601–612.
- Bach, K., Szczepanski, T., Aamodt, A., Gundersen, O. E., & Mork, P. J. (2016). Case representation and similarity assessment in the selfBACK decision support system. In *International conference on case-based* (pp. 32–46). Springer.
- Bair, M. J., Matthias, M. S., Nyland, K. A., Huffman, M. A., Stubbs, D. L., Kroenke, K., & Damush, T. M. (2009). Barriers and facilitators to chronic pain self-management: A qualitative study of primary care patients with comorbid musculoskeletal pain and depression. *Pain Medicine*, 10, 1280–1290.
- Bernstein, I. A., Malik, Q., Carville, S., & Ward, S. (2017). Low back pain and sciatica: Summary of NICE guidance. *British Medical Journal*, 356, i6748.
- Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L. O. P., Foster, N. E., Grotle, M., Koes, B. W., Kovacs, F. M., Lin, C. C., Maher, C. G., Pearson, A. M., Peul, W. C., Schoene, M. L., Turk, D. C., van Tulder, M. W., Terwee, C. B., & Ostelo, R. W. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*, 159, 481–495.
- Chiarotto, A., Maxwell, L. J., Terwee, C. B., Wells, G. A., Tugwell, P., & Ostelo, R. W. (2016). Roland-Morris disability questionnaire and Oswestry disability index: Which has better measurement properties for measuring physical functioning in nonspecific low Back pain?. Systematic Review and Meta-Analysis. *Physical Therapy*, 96, 1620–1637.
- Clauw, D. J., Essex, M. N., Pitman, V., & Jones, K. D. (2019). Reframing chronic pain as a disease, not a symptom: Rationale

- and implications for pain management. *Postgraduate Medicine*, 131, 185–198.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.
- Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health: Claremont symposium on applied social psychology*. Sage.
- Currie, S. R., & Wang, J. (2005). More data on major depression as an antecedent risk factor for first onset of chronic back pain. *Psychological Medicine*, 35, 1275–1282.
- Dennison, L., Morrison, L., Conway, G., & Yardley, L. (2013). Opportunities and challenges for smartphone applications in supporting health behavior change: Qualitative study. *Journal of Medical Internet Research*, 15, e86.
- Dunn, K. M., Jordan, K., & Croft, P. R. (2006). Characterizing the course of low back pain: A latent class analysis. *American Journal of Epidemiology*, 163, 754–761.
- Equipsme. (2022). Patients waiting up to four months to see a physiotherapist in England. In *The Know*. Equipsme. <https://www.equipsme.com/blog/up-to-four-months-to-see-a-physiotherapist/>
- Foster, N. E., Anema, J. R., Cherkin, D., Chou, R., Cohen, S. P., Gross, D. P., Ferreira, P. H., Fritz, J. M., Koes, B. W., & Peul, W. (2018). Prevention and treatment of low back pain: Evidence, challenges, and promising directions. *Lancet*, 391, 2368–2383.
- Grimby, G., Börjesson, M., Jonsdottir, I., Schnohr, P., Thelle, D., & Saltin, B. (2015). The “Saltin-Grimby physical activity level scale” and its application to health research. *Scandinavian Journal of Medicine & Science in Sports*, 25, 119–125.
- Hartvigsen, J., Hancock, M. J., Kongsted, A., Louw, Q., Ferreira, M. L., Genevay, S., Hoy, D., Karpainen, J., Pransky, G., & Sieper, J. (2018). What low back pain is and why we need to pay attention. *Lancet*, 391, 2356–2367.
- Hill, J. C., Whitehurst, D. G., Lewis, M., Bryan, S., Dunn, K. M., Foster, N. E., Konstantinou, K., Main, C. J., Mason, E., Somerville, S., Sowden, G., Vohora, K., & Hay, E. M. (2011). Comparison of stratified primary care management for low Back pain with current best practice (STarT Back): A randomised controlled trial. *Lancet*, 378, 1560–1571.
- Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., & Barendregt, J. (2014). The global burden of low back pain: Estimates from the global burden of disease 2010 study. *Annals of the Rheumatic Diseases*, 73, 968–974.
- Jackson, T., Wang, Y., Wang, Y., & Fan, H. (2014). Self-efficacy and chronic pain outcomes: A meta-analytic review. *The Journal of Pain*, 15, 800–814.
- Jenks, A., Hoekstra, T., van Tulder, M., Ostelo, R. W., Rubinstein, S. M., & Chiarotto, A. (2022). Roland-Morris disability questionnaire, Oswestry disability index, and Quebec Back pain disability scale: Which has superior measurement properties in older adults with low Back pain? *The Journal of Orthopaedic and Sports Physical Therapy*, 52, 457–469.
- Jordan, K., Dunn, K. M., Lewis, M., & Croft, P. (2006). A minimal clinically important difference was derived for the Roland-Morris disability questionnaire for low back pain. *Journal of Clinical Epidemiology*, 59, 45–52.
- Kamper, S. J., Maher, C. G., & Mackay, G. (2009). Global rating of change scales: A review of strengths and weaknesses and considerations for design. *The Journal of Manual & Manipulative Therapy*, 17, 163–170.
- Kent, P., Grotle, M., Dunn, K. M., Albert, H. B., & Lauridsen, H. H. (2015). Rasch analysis of the 23-item version of the Roland Morris disability questionnaire. *Journal of Rehabilitation Medicine*, 47, 356–364.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, 114, 163–173.
- Lauridsen, H. H., Hartvigsen, J., Manniche, C., Korsholm, L., & Grunnet-Nilsson, N. (2006). Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskeletal Disorders*, 7, 82.
- Lee, E.-H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nursing Research*, 6, 121–127.
- Linardon, J., & Fuller-Tyszkiewicz, M. (2020). Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of Consulting and Clinical Psychology*, 88, 1–13.
- Mork, P. J., & Bach, K. (2018). A decision support system to enhance self-management of low Back pain: Protocol for the selfBACK project. *JMIR Res Protoc*, 7, e167.
- Nicholas, M. K. (2007). The pain self-efficacy questionnaire: Taking pain into account. *European Journal of Pain*, 11, 153–163.
- Ostelo, R. W., & de Vet, H. C. (2005). Clinically important outcomes in low back pain. *Best Practice & Research. Clinical Rheumatology*, 19, 593–607.
- Pincus, T., & McCracken, L. M. (2013). Psychological factors and treatment opportunities in low back pain. *Best Practice & Research. Clinical Rheumatology*, 27, 625–635.
- Pinheiro, M. B., Ferreira, M. L., Refshauge, K., Maher, C. G., Ordoñana, J. R., Andrade, T. B., Tsathas, A., & Ferreira, P. H. (2016). Symptoms of depression as a prognostic factor for low back pain: A systematic review. *The Spine Journal*, 16, 105–116.
- Power, C., Frank, J., Hertzman, C., Schierhout, G., & Li, L. (2001). Predictors of low back pain onset in a prospective British study. *American Journal of Public Health*, 91, 1671–1678.
- Puschmann, A.-K., Drießlein, D., Beck, H., Arampatzis, A., Catalá, M. M., Schiltenswolf, M., Mayer, F., & Wippert, P.-M. (2020). Stress and self-efficacy as long-term predictors for chronic low back pain: A prospective longitudinal study. *Journal of Pain Research*, 13, 613–621.
- Roland, M., & Fairbank, J. (2000). The Roland–Morris disability questionnaire and the Oswestry disability questionnaire. *Spine*, 25, 3115–3124.
- Rusu, A. C., Gajjar, H., Schlüter, M. C., & Bremer, Y. I. (2019). Cognitive biases toward pain: Implications for a neurocognitive processing perspective in chronic pain and its interaction with depression. *The Clinical Journal of Pain*, 35, 252–260.
- Rusu, A. C., & Hallner, D. (2018). Idiographic measurement of depressive thinking: Development and preliminary validation of the sentence completion test for chronic pain (SCP). *Scandinavian Journal of Pain*, 18, 491–503.

- Rusu, A. C., & Pincus, T. (2017). Chronic pain patients' perceptions of their future: A verbal fluency task. *Pain, 158*, 171–178.
- Sandal, L. F., Bach, K., Øverås, C. K., Svendsen, M. J., Dalager, T., Jensen, J. S. D., Kongsvold, A., Nordstoga, A. L., Bardal, E. M., & Ashikhmin, I. (2021). Effectiveness of app-delivered, Tailored self-management support for adults with lower Back pain-related disability: A selfBACK randomized clinical trial. *JAMA Internal Medicine, 181*(10), 1288–1296.
- Sandal, L. F., Stochkendahl, M. J., Svendsen, M. J., Wood, K., Øverås, C. K., Nordstoga, A. L., Villumsen, M., Rasmussen, C. D. N., Nicholl, B., & Cooper, K. (2019). An app-delivered self-management program for people with low back pain: Protocol for the selfBACK randomized controlled trial. *JMIR Research Protocols, 8*, e14720.
- Schäfer, I., Kaduszkiewicz, H., Wagner, H.-O., Schön, G., Scherer, M., & van den Bussche, H. (2014). Reducing complexity: A visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health, 14*, 1–14.
- Schubart, J. R., Stuckey, H. L., Ganeshamoorthy, A., & Sciamanna, C. N. (2011). Chronic health conditions and internet behavioral interventions: A review of factors to enhance user engagement. *Computers, Informatics, Nursing, 29*, 81–92.
- Stubbs, B., Koyanagi, A., Thompson, T., Veronese, N., Carvalho, A. F., Solomi, M., Mugisha, J., Schofield, P., Cosco, T., & Wilson, N. (2016). The epidemiology of back pain and its relationship with depression, psychosis, anxiety, sleep disturbances, and stress sensitivity: Data from 43 low-and middle-income countries. *General Hospital Psychiatry, 43*, 63–70.
- Wertli, M. M., Eugster, R., Held, U., Steurer, J., Kofmehl, R., & Weiser, S. (2014). Catastrophizing—A prognostic factor for outcome in patients with low back pain: A systematic review. *The Spine Journal, 14*, 2639–2657.
- Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., Rice, D. B., Boruff, J., Cuijpers, P., Gilbody, S., Ioannidis, J. P. A., Kloda, L. A., McMillan, D., Patten, S. B., Shrier, I., Ziegelstein, R. C., Akena, D. H., Arroll, B., Ayalon, L., ... Thombs, B. D. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: A systematic review and individual participant data meta-analysis. *Psychological Medicine, 50*, 1368–1380.
- Zhao, J., Freeman, B., & Li, M. (2016). Can mobile phone apps influence people's health behavior change? An evidence review. *Journal of Medical Internet Research, 18*, e287.

## SUPPORTING INFORMATION

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

**How to cite this article:** Rughani, G., Nilsen, T. I. L., Wood, K., Mair, F. S., Hartvigsen, J., Mork, P. J., & Nicholl, B. I. (2023). The selfBACK artificial intelligence-based smartphone app can improve low back pain outcome even in patients with high levels of depression or stress. *European Journal of Pain, 27*, 568–579. <https://doi.org/10.1002/ejp.2080>