

JAMA Internal Medicine | [Original Investigation](#) | [LESS IS MORE](#)

Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools

A Randomized Comparative Effectiveness Trial

John D. Piette, MSc, PhD; Sean Newman, MS; Sarah L. Krein, PhD; Nicole Marinac, MPH; Jenny Chen, MPH; David A. Williams, PhD; Sara N. Edmond, PhD; Mary Driscoll, PhD; Kathryn M. LaChappelle, MPH; Robert D. Kerns, PhD; Marianna Maly, MA; H. Myra Kim, ScD; Karen B. Farris, PhD; Diana M. Higgins, PhD; Eugenia Buta, PhD; Alicia A. Heapy, PhD

IMPORTANCE Cognitive behavioral therapy for chronic pain (CBT-CP) is a safe and effective alternative to opioid analgesics. Because CBT-CP requires multiple sessions and therapists are scarce, many patients have limited access or fail to complete treatment.

OBJECTIVES To determine if a CBT-CP program that personalizes patient treatment using reinforcement learning, a field of artificial intelligence (AI), and interactive voice response (IVR) calls is noninferior to standard telephone CBT-CP and saves therapist time.

DESIGN, SETTING, AND PARTICIPANTS This was a randomized noninferiority, comparative effectiveness trial including 278 patients with chronic back pain from the Department of Veterans Affairs health system (recruitment and data collection from July 11, 2017-April 9, 2020). More patients were randomized to the AI-CBT-CP group than to the control (1.4:1) to maximize the system's ability to learn from patient interactions.

INTERVENTIONS All patients received 10 weeks of CBT-CP. For the AI-CBT-CP group, patient feedback via daily IVR calls was used by the AI engine to make weekly recommendations for either a 45-minute or 15-minute therapist-delivered telephone session or an individualized IVR-delivered therapist message. Patients in the comparison group were offered 10 therapist-delivered telephone CBT-CP sessions (45 minutes/session).

MAIN OUTCOMES AND MEASURES The primary outcome was the Roland Morris Disability Questionnaire (RMDQ; range 0-24), measured at 3 months (primary end point) and 6 months. Secondary outcomes included pain intensity and pain interference. Consensus guidelines were used to identify clinically meaningful improvements for responder analyses (eg, a 30% improvement in RMDQ scores and pain intensity). Data analyses were performed from April 2021 to May 2022.

RESULTS The study population included 278 patients (mean [SD] age, 63.9 [12.2] years; 248 [89.2%] men; 225 [81.8%] White individuals). The 3-month mean RMDQ score difference between AI-CBT-CP and standard CBT-CP was -0.72 points (95% CI, -2.06 to 0.62) and the 6-month difference was -1.24 (95% CI, -2.48 to 0); noninferiority criterion were met at both the 3- and 6-month end points ($P < .001$ for both). A greater proportion of patients receiving AI-CBT-CP had clinically meaningful improvements at 6 months as indicated by RMDQ (37% vs 19%; $P = .01$) and pain intensity scores (29% vs 17%; $P = .03$). There were no significant differences in secondary outcomes. Pain therapy using AI-CBT-CP required less than half of the therapist time as standard CBT-CP.

CONCLUSIONS AND RELEVANCE The findings of this randomized comparative effectiveness trial indicated that AI-CBT-CP was noninferior to therapist-delivered telephone CBT-CP and required substantially less therapist time. Interventions like AI-CBT-CP could allow many more patients to be served effectively by CBT-CP programs using the same number of therapists.

TRIAL REGISTRATION ClinicalTrials.gov Identifier: [NCT02464449](#)

JAMA Intern Med. doi:10.1001/jamainternmed.2022.3178
Published online August 8, 2022.

[+ Visual Abstract](#)

[+ Supplemental content](#)

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: John D. Piette, MSc, PhD, Center for Managing Chronic Disease, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109 (jpiette@umich.edu).

Chronic pain is a prevalent and increasing problem¹⁻³ that has been associated with work interruption, emotional distress, and risky behaviors, including substance use.⁴ Treatment has emphasized pharmacotherapy and surgery, both of which are potentially costly options with variable efficacy and substantial risks.⁵⁻⁷ Cognitive behavioral therapy for chronic pain (CBT-CP) is a safe, evidence-based alternative with moderate to large effects on outcomes.⁸⁻¹¹ Typically delivered by a therapist during 6-12 weekly in-person sessions, CBT-CP targets maladaptive thought processes and promotes adaptive behaviors. Because therapists are scarce, multiple in-person sessions are burdensome, and reimbursement is limited, many patients do not have easy access to CBT-CP or receive a lower dose than intended.¹² Given ongoing concern about the epidemic of opioid-related harms, broader access to CBT-CP is vitally important.¹³⁻¹⁵

To improve access to CBT-CP, Heapy and colleagues¹⁶ developed the Cooperative Pain Education and Self-Management (COPEs) intervention using individually tailored interactive voice response (IVR) calls to deliver CBT-CP content and therapist feedback. A randomized trial found that COPEs achieved outcomes that were noninferior to standard in-person therapist-delivered CBT-CP. Although that trial supports the use of IVR-based CBT-CP, the ideal duration and mode of CBT-CP sessions remains unknown, and programs vary substantially.¹⁷ Patients with complex issues may benefit from longer sessions, others may be motivated by brief, live synchronous sessions, and still others may respond better to the less burdensome delivery of support afforded by IVR.¹⁸

As a follow-up to COPEs, we developed a CBT-CP intervention using artificial intelligence (AI) to automatically adjust the modality of weekly therapist interactions based on patient feedback reported daily via IVR. We evaluated that intervention (AI-CBT-CP) relative to therapist-delivered telephone CBT-CP in the REACT (Responsive, Efficient, Accessible Chronic Pain Technology) comparative effectiveness trial.¹⁹ In REACT, weekly recommendations were made by AI-CBT-CP regarding the mode and duration of therapist-patient interactions using reinforcement learning—a type of AI commonly used in robotics in which an intelligent agent learns to progressively refine decisions based on probabilistic trials of new choices coupled with feedback about the response. Reinforcement learning algorithms similar to those applied in REACT are the basis of recommendations on platforms such as Netflix and Amazon.com.²⁰ We hypothesized that AI-CBT-CP would produce improvements in pain-relevant outcomes that were not meaningfully inferior to therapist-delivered telephone CBT-CP.

Methods

The research protocol was approved by the Human Subjects Committees in 2 Department of Veterans Affairs (VA) medical centers. All patients provided written informed consent. The trial was registered and the Consolidated Standards of Reporting Trials (CONSORT) reporting guidelines were used to identify

Key Points

Question Can a cognitive behavioral therapy intervention for chronic pain (CBT-CP) that adjusts treatment using artificial intelligence (AI-CBT-CP) based on feedback about patient progress achieve outcomes that are not inferior to standard telephone CBT-CP while reducing therapist time?

Findings This randomized comparative effectiveness trial of AI-CBT-CP found that its outcomes were not inferior to those of 45-minute telephone therapist sessions, with less than half the therapist time. At 6 months, more patients who experienced AI-CBT-CP had clinically meaningful improvements in physical function and pain intensity.

Meaning The findings of this randomized trial indicated that AI-CBT-CP can achieve noninferior and possibly better outcomes relative to standard CBT-CP while increasing access and reducing therapist costs.

key information for reporting trial results. The trial protocol is available in [Supplement 1](#).

Patient Recruitment

Patients with chronic back pain were identified from the medical records of 2 VA health care systems from June 2017 to September 2019 using codes from the *International Classification of Diseases, Ninth Revision (ICD-9)* and the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)*. Eligible patients had 2 or more documented reports of moderate or greater pain intensity on the Numerical Rating Scale (NRS; score, ≥ 4) during the prior year²¹; moderate or greater pain-related disability using the Roland Morris Disability Questionnaire (RMDQ; score, ≥ 5)²²; moderate or greater musculoskeletal pain during 3 or more of the prior 6 months²¹; self-reported ability to walk at least 1 block; no symptoms or diagnoses of suicidality, severe depression,²³ uncontrolled psychosis, bipolar disorder, active substance use disorder, dementia, a disabling sensory deficit, or a life-threatening medical condition (eg, end-stage heart failure); a touch-tone telephone; no planned surgical intervention for pain; and no current participation in CBT-CP.

Randomization

After completing baseline assessments, patients were randomized to AI-CBT-CP or therapist-delivered telephone CBT-CP, stratified by recruitment site. Allocation was concealed using sequentially numbered, sealed envelopes containing computer-generated treatment assignments. To maximize the ability of AI-CBT-CP to learn from patient feedback, patients were randomized disproportionately into the AI-CBT-CP group (ratio, 1.4:1.0).

CBT-CP Elements in Both Groups

Both groups received a CBT-CP manual describing 10 weekly modules addressing 8 pain coping skills. All patients received an Omron HJ-320 pedometer (Omron Healthcare) for monitoring step counts.²⁴ During each session, patients chose a behavioral goal and received a daily walking goal of 110% of the

prior week's average steps. Therapists were master's- or PhD-level clinicians trained by a PhD-level or advanced practice registered nurse-level CBT-CP therapist. Therapists used a treatment manual and attended weekly group supervision to prevent intervention drift. A PhD-level researcher reviewed a sample of audiotaped sessions and rated treatment fidelity.²⁵

Therapist-Delivered Telephone CBT-CP (Comparison Group)

Patients in the comparison group were offered 10 weekly, 45-minute therapist-delivered telephone CBT-CP sessions. Sessions included a review of the patient's pedometer logs and skill practice, a presentation of new skill information, the selection of behavioral goals, and the use of problem-solving techniques to address goal completion barriers.

AI-CBT-CP Intervention

Session Recommendations or "Action Choices"

Each week, AI-CBT-CP recommended 1 of 3 options for each patient's session. Option 1 was an asynchronous (recorded) IVR-delivered session using content from the COPEs trial, which included a voice message with individualized therapist feedback based on the participant's IVR-reported data (further details to follow).¹⁶ Feedback messages acknowledged progress, identified connections between pain and other patient reports, and provided reinforcement. Option 2 was a 15-minute synchronous (live) telephone session in which the therapist provided reinforcement and addressed skill practice barriers. Patients who missed the session received IVR feedback. Option 3 was a 45-minute synchronous (live) telephone session that prioritized problem-solving difficulties with skill practice and progress toward physical activity goals. Patients could identify up to 3 problems to address; and encouragement, education, and problem solving were provided as needed. Patients who missed the session received IVR feedback.

Monitoring Patient Progress

The AI-CBT-CP session recommendations were based on information patients reported via brief (<5 minutes) daily IVR calls. During these calls, patients reported information about their step counts, sleep, pain intensity, interference, mood, self-efficacy, CBT skill practice, and progress toward behavioral goals. Patients who missed 2 or more consecutive calls were contacted to troubleshoot problems and encourage engagement. Patients who missed more than 4 calls in a given week were automatically assigned a 15-minute session. Patients were informed via IVR about the modality of their session for the week.

Evaluating Weekly Changes in Patient Status

To make session recommendations, AI-CBT-CP modified its probability distribution across the 3 possible session types or action choices, with the goal of optimizing the patient's future status. Patients' status each week was evaluated by the system based on a score composed in equal proportions of (1) the patient's step counts reported daily via IVR, and (2) the patient's experience of pain-related interference measured via 2 questions from the Brief Pain Inventory, also reported daily

via IVR.²⁶ Each week, AI-CBT-CP calculated the expected value of this score (the "reward" in AI) for each of the 3 action choices based on scores received after prior recommendations for that patient and others in the AI-CBT-CP group. If there was a tie in expected reward scores between more and less resource intensive action choices (eg, 45-minute session and IVR-delivered session), the expected reward scores associated with live therapist contacts were discounted using a cost factor of -0.02 for a 15-minute session and -0.06 for a 45-minute session.

Contextualized Learning

In calculating the expected reward associated with the 3 action choices, AI-CBT-CP used a multidimensional matrix of patient characteristics and experiences that allowed the system to personalize its recommendations. These variables reflected the patient's historical levels of physical activity, pain intensity, sleep, CBT skill practice, and the ordinal session number associated with a given reward-recommendation pair (eg, second session, ninth session).

Calculation

In calculating the dynamic relationship between action choices and expected rewards for each patient each week, AI-CBT-CP used the reinforcement learning algorithm *LinUCB*,²⁰ which is a nonparametric algorithm designed to adapt decisions rapidly in the context of sparse data. This is important because, unlike online adaptive systems that may receive millions of inputs per minute, AI-CBT-CP received feedback on less than a dozen parameters each week for a relatively small number of patients.

Measurement

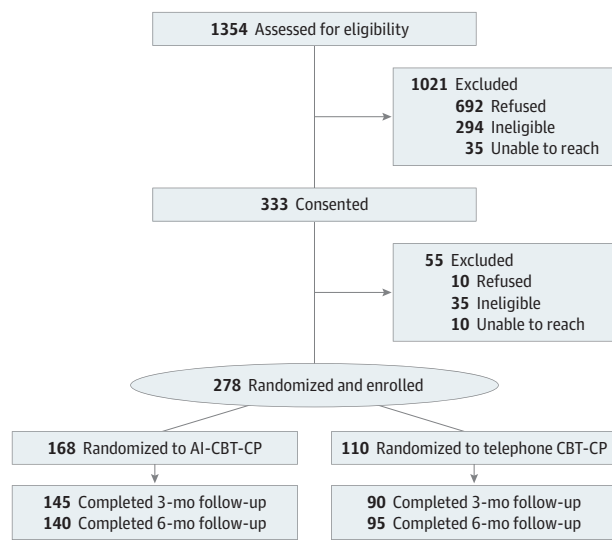
Primary Outcome

Outcomes were measured via telephone or mailed surveys at 3 months (primary outcome) and 6 months post baseline. Measures were selected based on expert recommendations from the IMMPACT trials (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials).^{21,27} The primary outcome for the trial was the RMDQ,^{21,27,28} a measure of pain-related disability for individuals with chronic noncancer pain.²²

Secondary Outcomes

Pain intensity was assessed using the pain NRS.²¹ Pain-related interference was measured using the Brief Pain Inventory interference items.^{21,26} Depression symptoms were assessed using the Patient Health Questionnaire.²³ We used the Veterans SF-12²⁹ to assess health-related quality of life within subdomains measuring patient physical functioning and mental functioning. Patients' overall impression of change since starting treatment was assessed using the Patient Global Impression of Change measure (PGIC).²¹ A widely adopted consensus statement and other experts in pain research recommend not only examining average changes in pain outcomes, but also the proportion of patients with clinically meaningful improvement in pain outcomes using standardized thresholds for defining improvement.^{21,30} Similar to the COPEs trial,¹⁶ we used the recommended 30% improvement to identify

Figure 1. CONSORT Diagram



responders regarding changes in RMDQ, NRS, and Brief Pain Inventory scores; and the recommended threshold of *at least moderately better* for the PGIC.²¹

Treatment Fidelity

A PhD-level CBT-CP expert ensured fidelity of treatment by rating a sample of audiotaped sessions for adherence and competence. A revised version of the Yale Adherence and Competence Scale was used to perform the rating.²⁵

Sociodemographic Information

Race and ethnicity were self-reported using standard US Census categories. Age, educational attainment, and other sociodemographic variables were also self-reported.

Sample Size and Power Calculation

The study was powered to detect noninferiority in RMDQ scores in the AI-CBT-CP group relative to the group receiving standard therapist-delivered telephone CBT-CP. Prior recommendations for minimally significant changes in RMDQ scores range from 2 to 8 points, and we selected the more stringent noninferiority margin of 2.^{22,31,32} To detect noninferiority with 90% power and type I error (1-sided) of 0.025, assuming an SD of 4.5 and an allocation ratio of 1.4:1, we needed 128 patients in the AI-CBT-CP group and 93 in the comparison group. To account for attrition, we recruited 278 patients.

Statistical Analysis

Baseline differences across groups in end point measures and other potential prognostic indicators were examined. The analytic cohort was intent-to-treat. A mixed model was used for repeated measurements to analyze RMDQ outcomes, including patients as random intercepts, and fixed effects of time (3-month, 6-month), treatment group, a treatment group-by-time interaction, the baseline value of the outcome, and recruitment site. Mixed models included all patients with avail-

able outcome data (ie, a patient was included in the analysis if they had an outcome at 3 months, even if they were missing at 6 months, and vice versa). Sensitivity analysis using multiple imputation for missing data that included all patients showed substantively the same results (available on request). Because age can be an important predictor of RMDQ scores^{33,34} and randomization groups had marginally different mean (SD) ages at baseline—62.9 (13.1) years among the AI-CBT-CP group vs 65.5 (10.6) years among the CBT-CP group (difference, 2.6 years; 95% CI, -0.1 to 5.8; $P = .06$)—age was included as a covariate in outcome analyses. Secondary continuous outcomes were analyzed using similar mixed models. Responder analyses of binary outcomes used generalized linear mixed models with a logit link, time, treatment group, time-by-treatment interaction, baseline values of the outcome, site, and age as fixed effects; and patients as random intercepts. The PGIC model did not include baseline values because baseline values are not applicable to the measure.

Statistical tests were 2-tailed and P values $< .05$ were considered to be statistically significant. Data analyses were performed from April 26, 2021, to May 18, 2022, using SAS, version 9.4 (SAS Institute Inc) and Stata, version 16.1 (StataCorp LLC).

Results

The study population comprised 278 patients (mean [SD] age, 63.9 [12.2] years; 248 [89.2%] men and 21 [12.5%] women; 36 [13.1%] Black, 225 [81.8%] White, and 14 [5.0%] individuals who self-reported being of other race or multiracial). Of the 278 total patients, 168 were randomized to the AI-CBT-CP group and 110 to the comparison group. Follow-up was high for both groups (Figure 1): 235 of 278 patients (85%) provided outcomes at the 3-month follow-up; 235 (85%) at the 6-month follow-up; and 251 (90%) provided outcomes for 1 or more end points. Other than the marginal difference in mean age, there were no notable baseline differences in participant characteristics between groups (Table 1). Patients reported several risk factors for poor health outcomes; more than half reported 1 or more emergency department visits in the prior year and 28.6% were hospitalized.

Engagement in Treatment and IVR Call Completion

Patients receiving AI-CBT-CP completed more treatment weeks than those in the comparison group (Mann-Whitney rank sum comparison, $P < .001$; Figure 2); 81.5% completed all 10 modules compared with 57.3% of comparison-group patients. Patients in the AI-CBT-CP group completed 9220 of 10 504 (87.8%) IVR calls. The AI-CBT-CP intervention made recommendations for weeks in which patients completed 3 or more IVR calls, and this was possible 94.2% of the time (1436 of 1525 treatment weeks). Patients received the session type recommended (ie, without substitution) 80.3% of the time. More details about AI-CBT-CP patient engagement with the intervention have been published.³⁵

A total of 45.8% of AI-CBT-CP sessions were delivered via IVR, 41.6% were 15-minute synchronous (live) sessions, and

Table 1. Baseline Characteristics of Patients, by Randomization Group

Characteristic	Randomization group		Standardized difference ^a
	AI-CBT-CP, No. (%)	Telephone CBT-CP, No. (%)	
Total patients	168	110	NA
Age, mean (SD), y	62.9 (13.1)	65.5 (10.6)	0.22
Sex			
Male	147 (87.5)	101 (91.8)	0.14
Female	21 (12.5)	9 (8.2)	
Race			
Black	26 (15.7)	10 (9.2)	0.21
White	133 (80.1)	92 (84.4)	
Other ^b	7 (4.2)	7 (6.4)	
Hispanic ethnicity	6 (3.6)	5 (4.6)	0.05
Married/partnered	108 (64.7)	70 (64.2)	0.01
Education			
High school or less	29 (17.4)	27 (24.8)	0.18
Some college	138 (82.6)	82 (75.2)	
Employment status			
Employed	50 (30.5)	34 (31.8)	0.03
Not employed	114 (69.5)	73 (68.2)	
Distance to VA, miles			
<10	20 (11.9)	23 (20.9)	0.24
10-19	36 (21.4)	21 (19.1)	
≥20	112 (66.7)	66 (60.0)	
Outpatient visits in past year			
≤5	52 (31.1)	27 (24.8)	0.19
6-10	39 (23.4)	34 (31.2)	
11-20	37 (22.2)	23 (21.1)	
>20	39 (23.4)	25 (22.9)	
ED visits in past year			
0 visits	74 (44.3)	52 (48.1)	0.20
1	40 (24.0)	31 (28.7)	
≥2	53 (31.7)	25 (23.1)	
Admitted in past year			
0 admissions	114 (68.3)	83 (76.1)	0.23
1	43 (25.7)	18 (16.5)	
≥2	10 (6.0)	8 (7.3)	
Missed physician visit in past year			
0 visits	97 (58.1)	71 (65.7)	0.22
1	39 (23.4)	16 (14.8)	
≥2	31 (18.6)	21 (19.4)	

Abbreviations: AI, artificial intelligence; CBT-CP, cognitive behavioral therapy intervention for chronic pain; ED, emergency department; VA, Department of Veterans Affairs health facility.

^a Differences in means or proportions between groups, divided by a pooled estimate of the standard deviation.

^b Respondents who self-reported more than 1 race; none reported being American Indian or Alaska Native, Asian, or Native Hawaiian or other Pacific Islander.

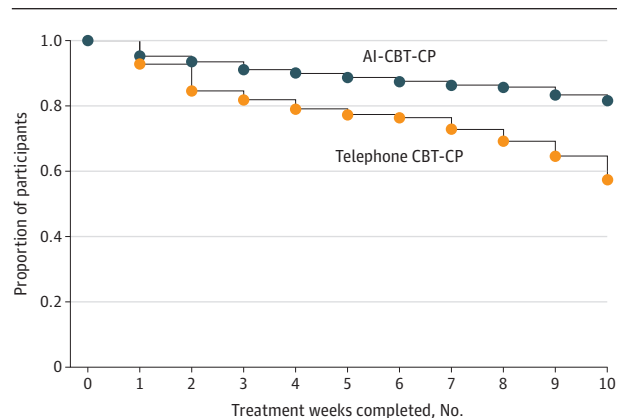
12.6% were 45-minute synchronous sessions. Compared with the standard CBT-CP intervention, the use of IVR and brief therapist contacts among the AI-CBT-CP group translated into a substantial reduction in therapist time. Assuming patients in both groups completed all 10 prescribed sessions, with the given distribution of session types, the AI-CBT-CP group would have received 26% as much synchronous therapist contact as controls, ie, 119 minutes per patient vs 450 minutes for controls. Even considering the greater number of contact weeks among patients receiving AI-CBT-CP, total live therapist time was only 30% as much as standard CBT-CP (ie, 111 minutes in an average of 9.3 sessions vs 365 minutes in an average of 8.1 sessions among controls). If asynchronous therapist

effort is included (ie, approximately 15 minutes preparing and recording each IVR message), AI-CBT-CP therapist time was 48% of that needed for standard CBT-CP.

Changes in Primary and Secondary Outcomes

At 3 months (primary end point), the mean RMDQ score was 10.95 (SE, 0.42) in the AI-CBT-CP group and 11.66 (SE, 0.53) in the comparison group (Table 2). The between-group difference was -0.72 points (95% CI, -2.06 to 0.62; $P < .001$ for non-inferiority), and because the upper limit of the confidence interval was below the noninferiority margin of 2.0 (), we concluded that AI-CBT-CP was noninferior to therapist-delivered telephone CBT-CP intervention. At 6 months, the dif-

Figure 2. Number of Completed Treatment Sessions, by Randomization Group



Abbreviations: AI, artificial intelligence; CBT-CP, cognitive behavioral therapy for chronic pain.

ference between AI-CBT-CP and the comparison group indicated that AI-CBT-CP was not only noninferior but marginally statistically superior (mean difference, -1.24 ; 95% CI, -2.48 to 0 ; $P = .05$). There was no evidence of a difference between groups regarding any secondary outcome at either assessment time.

Responder Analysis

At 3 months, there was no significant difference between groups in the proportion of patients with a clinically meaningful improvement in scores for any of the primary or secondary outcome measures (Table 3). At 6 months, a higher proportion of patients in the AI-CBT-CP group had a clinically meaningful improvement in RMDQ scores relative to comparison patients (37% vs 19%; $P = .01$; number needed to treat [NNT] = 6). Also at 6 months, a greater proportion of patients receiving AI-CBT-CP reported a clinically meaningful improvement in NRS scores (29% vs 17%; $P = .03$; NNT = 9). Differences in the proportion of patients reporting improvement in

Table 2. Primary and Secondary Outcomes (Mean Values) at Baseline, 3 Months, and 6 Months, and Differences Between Groups

Scale	AI-CBT-CP (n = 168)		Telephone CBT-CP (n = 110)		AI- vs telephone CBT-CP, mean (95% CI)	P value ^a
	No.	Mean (SE)	No.	Mean (SE)		
RMDQ^b						
Baseline	153	13.48 (0.33)	98	13.37 (0.43)	NA	NA
3 mo	145	10.95 (0.42)	90	11.66 (0.53)	$-0.72 (-2.06 \text{ to } 0.62)$	<.001 (noninferiority)
6 mo	139	11.26 (0.40)	95	12.50 (0.49)	$-1.24 (-2.48 \text{ to } 0)$	<.001 (noninferiority)
Pain Numerical Rating Scale^c						
Baseline	152	6.16 (0.12)	97	6.25 (0.15)	NA	NA
3 mo	142	5.04 (0.15)	89	5.13 (0.19)	$-0.09 (-0.59 \text{ to } 0.40)$.71
6 mo	138	5.34 (0.15)	94	5.70 (0.19)	$-0.36 (-0.84 \text{ to } 0.11)$.13
Brief Pain Inventory Interference^d						
Baseline	152	4.72 (0.17)	98	4.92 (0.21)	NA	NA
3 mo	145	3.92 (0.16)	89	3.93 (0.20)	$-0.01 (-0.52 \text{ to } 0.50)$.98
6 mo	138	4.04 (0.16)	93	4.45 (0.20)	$-0.41 (-0.91 \text{ to } 0.09)$.11
PHQ-9 Depression Scale^e						
Baseline	153	6.23 (0.42)	98	6.42 (0.51)	NA	NA
3 mo	145	6.72 (0.38)	90	7.23 (0.48)	$-0.51 (-1.71 \text{ to } 0.69)$.40
6 mo	138	6.88 (0.38)	94	7.38 (0.46)	$-0.50 (-1.68 \text{ to } 0.67)$.40
Physical Composite summary^f						
Baseline	153	27.28 (0.67)	98	27.23 (0.91)	NA	NA
3 mo	145	30.49 (0.65)	90	30.68 (0.82)	$-0.19 (-2.24 \text{ to } 1.87)$.86
6 mo	140	29.97 (0.68)	95	29.99 (0.84)	$-0.03 (-2.16 \text{ to } 2.10)$.98
Mental Composite summary^g						
Baseline	153	46.97 (1.01)	98	46.25 (1.28)	NA	NA
3 mo	145	49.04 (0.72)	90	48.10 (0.91)	$0.94 (-1.35 \text{ to } 3.23)$.42
6 mo	140	48.54 (0.73)	95	46.72 (0.89)	$1.82 (-0.45 \text{ to } 4.09)$.12

Abbreviations: AI, artificial intelligence; CBT-CP, cognitive behavioral therapy for chronic pain; NA, not applicable; PHQ-9, Patient Health Questionnaire-9; RMDQ, Roland Morris Disability Questionnaire.

^a For the primary outcome (RMDQ), the P values are from noninferiority tests. For the other outcomes, the P values are from 2-sided difference tests. Means at 3 and 6 mo are least-squares means (adjusted means) from the mixed models. Means at baseline are raw means.

^b Scale, 0-24 (higher scores indicate greater physical disability).

^c Scale, 0-10 (higher scores indicate worse pain).

^d Scale, 0-10 (higher scores indicate greater interference).

^e Scale, 0-27, (higher scores indicate more depression symptoms).

^f Scale, 0-100 (higher scores indicate better function).

^g Scale, 0-100 (higher scores indicate better function).

Table 3. Clinically Significant Improvement in Outcomes From Responder Analyses

Scale	AI-CBT-CP (n = 168)		Telephone CBT-CP (n = 110)		Adjusted odds ratio (95% CI)	P value
	Total No.	No. (%)	Total No.	No. (%)		
RMDQ ^a						
3 mo	145	54 (37)	90	29 (32)	1.51 (0.49-4.65)	.47
6 mo ^b	139	51 (37)	95	18 (19)	4.71 (1.39-15.93)	.01
NRS ^c						
3 mo	142	38 (27)	89	29 (33)	0.78 (0.22-2.73)	.69
6 mo ^d	138	40 (29)	94	16 (17)	4.54 (1.14-18.10)	.03
BPI ^e						
3 mo	145	53 (37)	89	33 (37)	1.00 (0.37-2.68)	>.99
6 mo	138	46 (33)	93	27 (29)	1.35(0.49-3.70)	.56
PGIC ^f						
3 mo	142	65 (46)	88	43 (49)	0.84 (0.37-1.90)	.68
6 mo ^g	137	68 (50)	90	31 (34)	2.24 (0.97-5.16)	.06

Abbreviations: AI, artificial intelligence; BPI, Brief Pain Inventory; CBT-CP, cognitive behavioral therapy intervention for chronic pain; NNT, number needed to treat; NRS, Numerical Rating Scale; PGIC, Patient Global Impression of Change; RMDQ, Roland Morris Disability Questionnaire.

^a Clinically significant improvement defined as a 30% improvement in scores relative to baseline.

^b Absolute difference, 18%; NNT = 6 (95% CIs, 4-15).

^c Clinically significant improvement defined as a 30% improvement in scores

relative to baseline.

^d Absolute difference, 12%; NNT = 9 (95% CI, 4-81).

^e Clinically significant improvement defined as a 30% improvement in scores relative to baseline.

^f Clinically significant improvement defined as report of at least "moderately better" pain control relative to baseline.

^g Absolute difference, 16%; NNT = 7 (95% CI, 4-44).

PGIC at 6 months also favored AI-CBT-CP (50% vs 34%), although the difference did not reach statistical significance ($P = .06$).

Discussion

In this trial, patients with chronic back pain randomized to 10 weeks of AI-CBT-CP had noninferior outcomes for pain-related functioning at 3 months post baseline compared with patients randomized to 10 weeks of 45-minute telephone sessions with a CBT-CP therapist. The AI-CBT-CP group had a marginally statistically significant better average RMDQ score at 6 months; however, this difference between groups is below the threshold considered clinically meaningful. At 3 and 6 months, average scores for other outcomes did not differ significantly between groups.

Responder analyses provide an easily interpretable measure of intervention effectiveness for clinicians and policy makers and are a recommended secondary analysis in pain studies.^{16,21,36} Using recommended thresholds, at 6 months almost twice as many patients receiving AI-CBT-CP had a clinically meaningful improvement in the primary outcome measure (RMDQ) and in pain intensity. Also at 6 months, a greater proportion of the AI-CBT-CP group reported that they had experienced at least moderate improvements since starting treatment (50% vs 34%; $P = .06$). Given the distribution of session types, AI-CBT-CP achieved these outcomes with only 30% of the clinician time required for the comparison program of weekly 45-minute therapist sessions.

Patients in the AI-CBT-CP group received a larger treatment dose than patients in the comparison group (82% com-

pleted all weekly sessions vs 57%). This is consistent with findings from the COPEs trial in which IVR-delivered treatment demonstrated an advantage over in-person treatment regarding the number of sessions completed (8.9 vs 6.6).¹⁶ However, it is important to note that in the current trial, the increase in patient engagement was demonstrated although the comparison intervention was also delivered by telephone.

Like clinicians, AI-CBT-CP can only make effective decisions about treatment course if it has feedback from reliable and valid assessments about patient status over time. In this study, brief daily IVR calls were successful in obtaining this feedback; AI-CBT-CP had the data it needed to make a decision 94% of the time. Other methods of collecting patient feedback (eg, text messaging, smartphone applications, automatically uploading pedometer readings, environmental sensors) should be explored as alternative strategies for providing interventions such as AI-CBT-CP with the information they need to be effective.

In contrast to other applications of reinforcement learning for which AI systems can learn from millions of data points in short intervals, the AI engine in this intervention received data on only a relatively small number of patients and interactions. Consequently, the program's effectiveness in this trial may represent a lower bound of what could be expected if it were implemented in larger samples of patients over longer time periods. Secondary analysis of data from the AI-CBT-CP group suggest that the intervention increased its effectiveness as it gained experience through patient interactions.³⁵ Future studies should seek to maximize the experience of AI-CBT-CP and similar programs through trials with larger populations and quantify more precisely the influence of program learning on patient health status.

The AI algorithm used to drive decision-making in the current intervention reflected a large number of design features and decisions informed by a panel of CBT-CP experts and best practices in reinforcement learning.¹⁹ Some of these decisions were made with incomplete information, and different results may have been obtained with different features. For example, AI-CBT-CP was designed to discount rewards achieved through more human resource intensive options. The magnitude of those cost coefficients and other program features are subjective, and future research should evaluate their associations with system performance and outcomes.

Limitations

This trial had several limitations. The results of both interventions were relatively modest. Patients were recruited from the US VA health care system, and the ways in which the trial implementation and findings would be different if conducted in other settings is difficult to anticipate. In general, patients in the VA system tend to be older, White, and male, and have a high burden of comorbid illnesses, poor health-related quality of life, mental health disorders, and substance use problems. These factors may limit the extent to which a new mode of treatment such as AI-CBT-CP can improve pain outcomes.³⁷⁻³⁹ Future interventions should consider integrating this model of CBT-CP with a focus on comorbid conditions that can affect

pain outcomes, such as posttraumatic stress disorder, depression, and substance use disorders. Alternatively, patients in the VA are often highly compliant with expectations such as responding to IVR calls,^{16,40,41} and that modality of data collection for AI-CBT-CP feedback (and consequently, AI-CBT-CP's decisions) may be less successful in other contexts. As in most pain trials, outcomes were self-reported and it was not possible to blind patients to their randomization condition; consequently, outcomes may reflect reporting bias.

Conclusions

This randomized noninferiority comparative effectiveness trial indicated that despite using less therapist time, AI-CBT-CP achieved outcomes that were noninferior to outcomes of patients offered an equal number of 45-minute telephone sessions with a CBT-CP therapist. Responder analyses suggest that during 6 months, more patients may achieve clinically meaningful improvements in pain control with AI-CBT-CP than with standard CBT-CP approaches. Given that AI-CBT-CP required less clinician-patient contact time, patients may find the intervention more convenient, and health systems could use it to treat more patients without additional clinical resources.

ARTICLE INFORMATION

Accepted for Publication: June 9, 2022.

Published Online: August 8, 2022.

doi:10.1001/jamainternmed.2022.3178

Author Affiliations: Ann Arbor Department of Veterans Affairs (VA) Center for Clinical Management Research, Ann Arbor, Michigan (Piette, Newman, Krein, Marinec, Chen, Maly, Kim); School of Public Health, University of Michigan, Ann Arbor (Piette, Newman, Marinec, Chen, Maly, Kim); Department of Internal Medicine, School of Medicine, University of Michigan, Ann Arbor (Piette, Krein); Chronic Pain and Fatigue Research Center, University of Michigan, Ann Arbor (Williams); Pain Research, Informatics, Multimorbidities, and Education (PRIME) Center, VA Connecticut Healthcare System, West Haven, Connecticut (Edmond, Driscoll, LaChappelle, Kerns, Heapy); Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut (Edmond, Driscoll, Kerns, Heapy); Department of Psychology, Yale University, New Haven, Connecticut (Kerns); Department of Neurology, Yale School of Medicine, New Haven, Connecticut (Kerns); Department of Clinical, Social, and Administrative Sciences, College of Pharmacy, University of Michigan, Ann Arbor (Farris); VA Boston Healthcare System, Boston, Massachusetts (Higgins); Boston University School of Medicine, Boston, Massachusetts (Higgins); Yale Center for Analytical Sciences, Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut (Buta).

Author Contributions: Dr Piette had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Piette, Krein, Marinec, Williams, Edmond, Kerns, Farris, Higgins, Heapy.

Acquisition, analysis, or interpretation of data:

Piette, Newman, Chen, Edmond, Driscoll, LaChappelle, Kerns, Maly, Kim, Buta, Heapy.
Drafting of the manuscript: Piette, Newman, Marinec, Buta.

Critical revision of the manuscript for important intellectual content: Piette, Krein, Marinec, Chen, Williams, Edmond, Driscoll, LaChappelle, Kerns, Maly, Kim, Farris, Higgins, Buta, Heapy.
Statistical analysis: Piette, Newman, Kim, Buta.
Obtained funding: Piette, Farris, Heapy.

Administrative, technical, or material support: Piette, Newman, Marinec, Chen, Edmond, Driscoll, LaChappelle, Maly, Higgins, Heapy.
Supervision: Piette, Marinec, Williams, Edmond, Driscoll, Heapy.

Conflict of Interest Disclosures: Dr Krein reported grants from the Department of Veterans Affairs (VA; Nos. IIR 13-350, RCS 11-222) during the conduct of the study. Dr Williams reported personal fees from Swing Therapeutics and Community Health Focus outside the submitted work. Dr Farris reported grants from AstraZeneca outside the submitted work. Dr Higgins reported grants from VA Office of Research and Development during the conduct of the study. Dr Heapy reported grants from the VA Health Services Research 7 Development during the conduct of the study. No other disclosures were reported.

Funding/Support: This study was funded by VA Health Services Research & Development (No. IIR #13-350). Drs Piette and Krein are VA Research Career Scientists.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Information: Several of the authors (Drs Piette, Williams, Edmond, Driscoll, Kerns, Higgins, and Heapy) participated in the Expert Panel for the REACT trial. None of the panel members received financial remuneration. Additional Expert Panel members included Mark A. Ilgen, PhD (VA Center for Clinical Management Research and University of Michigan School of Medicine) and Satinder Singh, PhD (University of Michigan College of Engineering).

Data Sharing Statement: See Supplement 2.

REFERENCES

- Rice ASC, Smith BH, Blyth FM. Pain and the global burden of disease. *Pain*. 2016;157(4):791-796. doi:10.1097/j.pain.0000000000000454
- Zajacova A, Grol-Prokopczyk H, Zimmer Z. Pain trends among American adults, 2002-2018: Patterns, disparities, and correlates. *Demography*. 2021;58(2):711-738. doi:10.1215/00703370-8977691
- Croft P, Blyth FM, van der Windt D. Chronic pain as a topic for epidemiology and public health. In: Croft P, Blyth FM, van der Windt D, eds. *Chronic Pain Epidemiology: From Etiology to Public Health*. Oxford; 2011:3-8.
- Kerns RD, Otis J, Rosenberg R, Reid MC. Veterans' reports of pain and associations with ratings of health, health-risk behaviors, affective distress, and use of the healthcare system. *J Rehabil Res Dev*. 2003;40(5):371-379. doi:10.1682/JRRD.2003.09.0371
- Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain—United States, 2016. *JAMA*. 2016;315(15):1624-1645. doi:10.1001/jama.2016.1464
- Krebs EE, Gravelly A, Nugent S, et al. Effect of opioid vs nonopioid medications on pain-related

- function in patients with chronic back pain or hip or knee osteoarthritis pain: the SPACE randomized clinical trial. *JAMA*. 2018;319(9):872-882. doi:10.1001/jama.2018.0899
7. US Department of Defense, Department of Veterans Affairs. *Clinical Practice Guideline for Opioid Therapy for Chronic Pain, v 3.0*. Department of Defense; 2017.
 8. Chou R, Huffman LH; American Pain Society; American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med*. 2007;147(7):492-504. doi:10.7326/0003-4819-147-7-200710020-00007
 9. Turk DC, Meichenbaum D, Genest M. *Pain and Behavioral Medicine: A Cognitive Behavioral Perspective*. Guilford Press; 1983.
 10. Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. Meta-analysis of psychological interventions for chronic low back pain. *Health Psychol*. 2007;26(1):1-9. doi:10.1037/0278-6133.26.1.1
 11. Williams AC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database Syst Rev*. 2012;11:CD007407. doi:10.1002/14651858.CD007407.pub3
 12. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. *Am Psychol*. 2014;69(2):153-166. doi:10.1037/a0035747
 13. Bohnert ASB, Ilgen MA. Understanding links among opioid use, overdose, and suicide. *N Engl J Med*. 2019;380(1):71-79. doi:10.1056/NEJMr1802148
 14. US National Center for Health Statistics, Centers for Disease Control and Prevention. Drug Overdose Deaths in the US Top 100,000 Annually. Accessed May 8, 2022. https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm
 15. US Centers for Disease Control and Prevention. US Opioid Dispensing Rate Maps. Accessed May 15, 2022. <https://www.cdc.gov/drugoverdose/rxrate-maps/index.html>
 16. Heapy AA, Higgins DM, Goulet JL, et al. Interactive voice response-based self-management for chronic back pain: the COPES noninferiority randomized trial. *JAMA Intern Med*. 2017;177(6):765-773. doi:10.1001/jamainternmed.2017.0223
 17. Henschke N, Ostelo RW, van Tulder MW, et al. Behavioural treatment for chronic low-back pain. *Cochrane Database Syst Rev*. 2010;(7):CD002014.
 18. Lambert MJ. Early response in psychotherapy: further evidence for the importance of common factors rather than "placebo effects". *J Clin Psychol*. 2005;61(7):855-869. doi:10.1002/jclp.20130
 19. Piette JD, Krein SL, Striplin D, et al. Patient-centered pain care using artificial intelligence and mobile health tools: protocol for a study funded by the US Department of Veterans Affairs Health Services Research and Development Program. *JMIR Res Protoc*. 2016;5(2):e53. doi:10.2196/resprot.4995
 20. Li L, Chu W, Langford J, Schapire RE. A contextual-bandit approach to personalized news article recommendation. Proceedings of the 19th International Conference on World Wide Web. April 26-30, 2010:661-670.
 21. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain*. 2008;9(2):105-121. doi:10.1016/j.jpain.2007.09.005
 22. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976)*. 2000;25(24):3115-3124. doi:10.1097/00007632-200012150-00006
 23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x
 24. Park W, Lee VJ, Ku B, Tanaka H. Effect of walking speed and placement position interactions in determining the accuracy of various newer pedometers. *J Exerc Sci Fit*. 2014;12(1):31-37. doi:10.1016/j.jesf.2014.01.003
 25. Carroll KM, Nich C, Sifry RL, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug Alcohol Depend*. 2000;57(3):225-238. doi:10.1016/S0376-8716(99)00049-6
 26. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *Clin J Pain*. 2004;20(5):309-318. doi:10.1097/00002508-200409000-00005
 27. Dworkin RH, Turk DC, Farrar JT, et al; IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1-2):9-19. doi:10.1016/j.pain.2004.09.012
 28. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)*. 1983;8(2):141-144. doi:10.1097/00007632-198303000-00004
 29. Kazis LE, Lee A, Spiro A III, et al. Measurement comparisons of the medical outcomes study and veterans SF-36 health survey. *Health Care Financ Rev*. 2004;25(4):43-58.
 30. Henschke N, van Enst A, Froud R, Ostelo RW. Responder analyses in randomised controlled trials for chronic low back pain: an overview of currently used methods. *Eur Spine J*. 2014;23(4):772-778. doi:10.1007/s00586-013-3155-0
 31. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheumatol*. 2001;28(2):431-438.
 32. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine (Phila Pa 1976)*. 1995;20(17):1899-1908. doi:10.1097/00007632-199509000-00011
 33. Rundell SD, Sherman KJ, Heagerty PJ, et al. Predictors of persistent disability and back pain in older adults with a new episode of care for back pain. *Pain Med*. 2017;18(6):1049-1062.
 34. Wilkens P, Scheel IB, Grundnes O, Hellum C, Storheim K. Prognostic factors of prolonged disability in patients with chronic low back pain and lumbar degeneration in primary care: a cohort study. *Spine (Phila Pa 1976)*. 2013;38(1):65-74. doi:10.1097/BRS.0b013e318263bb7b
 35. Piette JD, Newman S, Krein SL, et al. Artificial intelligence (AI) to improve chronic pain care: evidence of AI learning. *Intelligence-Based Medicine*. 2022;6:100064. doi:10.1016/j.ibmed.2022.100064
 36. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. *Pain*. 2010;151(3):694-702. doi:10.1016/j.pain.2010.08.034
 37. Kerns RD, Burns JW, Shulman M, et al. Can we improve cognitive-behavioral therapy for chronic back pain treatment engagement and adherence? a controlled trial of tailored versus standard therapy. *Health Psychol*. 2014;33(9):938-947. doi:10.1037/a0034406
 38. Dobscha SK, Corson K, Perrin NA, et al. Collaborative care for chronic pain in primary care: a cluster randomized trial. *JAMA*. 2009;301(12):1242-1252. doi:10.1001/jama.2009.377
 39. Zulman DM, Asch SM, Martins SB, Kerr EA, Hoffman BB, Goldstein MK. Quality of care for patients with multiple chronic conditions: the role of comorbidity interrelatedness. *J Gen Intern Med*. 2014;29(3):529-537. doi:10.1007/s11606-013-2616-9
 40. Piette JD, Rosland AM, Marinec NS, Striplin D, Bernstein SJ, Silveira MJ. Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically ill patients. *Med Care*. 2013;51(3):216-223. doi:10.1097/MLR.0b013e318277ebf8
 41. Piette JD, Striplin D, Marinec N, et al. A mobile health intervention supporting heart failure patients and their informal caregivers: a randomized comparative effectiveness trial. *J Med Internet Res*. 2015;17(6):e142. doi:10.2196/jmir.4550