Cover page

Official title:	**Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools**
NCT number:	**NCT02464449**
Document date:	**December 14, 2018**

B. BACKGROUND

B.1 Prevalence and Consequences of Chronic Pain among Veterans

Musculoskeletal disorders are highly prevalent among VA patients, with chronic back pain the most frequently reported type.[7,8] Among OEF/OIF Veterans, back pain and other musculoskeletal conditions are the most prevalent of all diagnosed medical and psychiatric conditions.[9] VA data suggest an annualized increase in the prevalence of low back pain of 4.8% per year due to factors such as an aging population and increasing prevalence of obesity.[7,10] Cost of treating back pain in VA is $2.2 billion annually.[8] Chronic low back pain is associated with work interruption, emotional distress, and risky heath behaviors such as substance use.[11] Emerging evidence suggests that chronic pain compromises successful treatment and management of other chronic conditions.[12] For all of these reasons, increasing access to effective, convenient treatments for chronic low back pain is a national VA priority.[13]

B2. Limitations of Pharmacologic Interventions for Chronic Low Back Pain

Historically, treatment for chronic low back pain has emphasized pharmacotherapy and surgery, while under-utilizing evidence-based behavioral approaches that have comparable or superior benefit.[14] Opioid medications are commonly used to manage severe chronic pain, but their use can lead to serious adverse effects. Prescription opioid use has increased 300% nationally since 1999, which has led to a 3-fold increase in the number of prescription opioid overdose deaths and half a million emergency room visits in 2009 alone.[15] In a sample of 762 OEF/OIF Veterans treated in VA for chronic non-cancer pain, 64% were prescribed an opioid medication during the prior year, and 41% received chronic opioid therapy.[16] Despite its frequent use, there is no evidence for the long term efficacy of opioid therapy for chronic pain.[14]

B.3 Cognitive and Behavioral Interventions to Improve Pain Management

Cognitive Behavioral Therapy (CBT) is the most widely-accepted evidence-based psychological treatment for chronic pain.[17] CBT is informed by theory recognizing that patients' beliefs, attitudes and coping styles play central roles in determining their experiences of pain.[18] CBT is an attractive alternative to pharmacotherapy because impacts on functioning can last long after treatment is discontinued, and CBT does not entail the negative side effects of opiods. The goal of pain CBT is to assist patients in developing an adaptive problem solving approach to pain management, and CBT targets both reductions in pain symptoms as well as their associated disability and emotional distress. VA recommendations regarding pain CBT, endorsed by the VA Evidence-Based Psychotherapy (EBP) program, recommend 10 hour-long sessions delivered weekly. Sessions provide pain education, teach and encourage the practice pain self-management skills, and promote productive and pleasurable activity and exercise. Skills address both cognitive processes (e.g., catastrophizing) and behaviors (e.g., relaxation). Dr. Robert Kerns (study Co-Investigator) published a meta-analysis demonstrating moderate to large effects of CBT in improving pain-related outcomes,[19] and a Cochrane systematic review reached nearly identical conclusions.[20]

Less than half of Veterans with an opioid prescription receive any mental health treatment,[21] and a survey by VA's National Program for Pain Management found that half of VA facilities have no pain-focused CBT. Given the gap between available supply and demand, VA needs to identify creative strategies to ensure that patients receive the self-management support that they need, while avoiding ineffective targeting of scarce clinical resources among patients who could benefit equally from less labor intensive forms of assistance. For patients receiving pain CBT, assessments of patients' status vary substantially across: therapists, patients, and time. Retrospective patient reports are often collected using pencil and paper surveys, and are vulnerable to recall and social desirability biases, e.g., reports may be disproportionately influenced by recent experiences and patients' emotional state at the time of assessment. Patient feedback is least likely to be available among Veterans with the greatest risk for missing in-person sessions, i.e., the very patients who may have the greatest need for adjustments in their treatment plan. Pain CBT guidelines are vague regarding how feedback should be used to shape subsequent therapy sessions. For all of these reasons, scarce CBT services can be slow to adapt to variation in patients' treatment response.

New models of CBT need to incorporate a systematic stepped approach to ensure that care is patient-centered, efficient, and targeted to Veterans' unique needs. VA CBT pain treatment is based on evidence

that typically reflects average effects in controlled trials, rather than taking into account the substantial variation across patients in treatment response. Recommendations for 10 hour-long CBT pain treatment sessions likely represent the upper bound of what is feasible given VA budgets and some patients' limited tolerance for frequent treatment. As many as 25% of patients receiving psychotherapy improve after 1-2 sessions[22] and patients often drop out of treatment that provides little marginal benefit. Some evidence-based CBT programs have as few as 6 sessions while others have twice that many.[23] Given that more than 800,000 VA patients have low back pain,[7] the significance of a difference in resources required to reach all patients with 6 versus 12 sessions is clear. Providers vary substantially in their recommendations regarding the appropriate visit frequency for chronic disease management.[24,25] Physicians typically make these decisions based on factors such as the patient's perceived stability or likelihood of treatment response. However some clinicians are poor judges of those characteristics, and evidence is weak regarding what visit frequencies are "right" for managing chronic diseases. One study found no correlation between visit frequency and hypertension control,[26] and visit intervals can sometimes be substantially lengthened without decreasing quality.[27] No single "dose" of CBT is likely to be appropriate for all Veterans, and neither clinicians nor patients may be able to judge *a priori* who needs more resource-intensive forms of care.

B.4 Prior Research on Adapting Treatment to Patients' Individual Needs

Lambert and colleagues demonstrated the benefits of adapting psychotherapy based on feedback to therapists about patients' progress.[28-30] While these studies represent an important step toward the type of systematically stepped treatment that we will evaluate in the proposed study, prior efforts have used occasional patient surveys to obtain information about treatment response at only a few points. As a result, the opportunity for adjusting therapy has been limited, and the impact of the patient tailoring has been modest. Other investigators have suggested that monitoring and feedback could best be accomplished using health IT[31] (such as we propose in the current study) to allow treatment decisions to be based on real-time information about patients' functioning. Another key weakness of prior work is that feedback on treatment response is typically provided to clinicians along with non-evidence-based algorithms for modifying patients' treatment plan.[22] As such, important initial steps toward a more systematic and evidence-based approach to adaptive treatment have been left with a format that cannot respond effectively to real-time information about what works best for each patient.

Another foundational area of research for the proposed study is the theory of tailored health communication, which suggests that patients are more likely to internalize health messages when those messages are relevant to them personally.[32] The state-of-the-science in tailoring uses surveys to identify patients' needs, health beliefs, learning styles, cultural context, and other factors prior to crafting health related messages targeting behavioral changes. The data needed to tailor health messages is substantial, and many patients may not be willing or able to accurately report that information at program outset.[33] For example, Drs. Kerns and Heapy found that CBT skills training was no more effective when skill presentation was tailored according to what patients thought they wanted before initiating treatment.[3] Also, previous tailoring systems typically have tailored based on static patient traits, rather than on updated information about patients' status or treatment response. In the proposed study, we will tailor Veterans' pain CBT services using IVR-reported feedback about their: pain-related physical functioning measured objectively via pedometer step counts, perceived functioning scores, and progress with CBT skill practice. Based on this real-time feedback, AI-CBT will personalize each patient's course of treatment automatically in order to achieve the greatest benefits for the population while using clinical resources as efficiently as possible.

B.5 Mobile Health (mHealth) Approaches to Self-Management Support

Because mHealth services have low marginal costs, they can cost-effectively reach large numbers of patients between face-to-face encounters to provide self-management support.[34] More than 50 studies have demonstrated that patients can provide reliable and valid information about psychiatric symptoms and substance abuse disorders via IVR and other mobile health technology.[5,6,35,36] Trials suggest that mHealth interventions can improve self-management behaviors[37,38] and may improve outcomes of chronic illness care.[39] The benefits of standard CBT diminish after patients discontinue therapy, and maintenance interventions delivered via IVR sustain those improvements in symptoms and self-management skills.[39,40]

Because mHealth services are so promising, VA is making significant investments in development through the national Office of Telehealth.[41] Despite their potential, mHealth interventions typically deliver simplistic series of messages based on pre-determined "if-then" rules and deterministic protocols. As a result, mHealth interactions can feel "robotic" to users and many drop out of treatment.[42] By way of analogy, no patient would continue to see a physician if that provider failed to adapt their management approach or goals based on the patient's response to prior treatments. We propose a model for taking advantage of the cost and accessibility benefits of mHealth services, while ensuring that these powerful tools are integrated systematically with personal, professional care from trained VA CBT therapists.

B.6 Significance to Veterans and VHA, and Responsiveness to HSR&D Priorities

This study seeks to identify a scalable and cost-effective strategy for increasing the number of Veterans who have access to effective CBT management for chronic back pain. Because this approach will systematically use "live" telephone therapy plus IVR, the service will be much more accessible than in-person CBT, increasing access while decreasing VA's dependence on over-burdened staff. The proposed study addresses two HSR&D priority areas: Health Informatics and Access to Care for Rural Veterans. This use of telehealth technology could allow VA to provide CBT pain self-management support to much larger numbers of Veterans, despite transportation barriers or the limited number of VA CBT therapists. Telephone CBT has been shown to produce positive improvement in pain and pain-relevant variables,[43,44] and the AI system will ensure that Veterans receive the appropriate intensity of care to reach treatment goals. IVR is a communication modality with strong evidence for feasibility and effectiveness among Veterans as well as non-VA populations with mental health and substance abuse disorders.[41,45,46] If successful, this trial will provide a template for using telehealth and artificial intelligence to provide efficient, personalized services for Veterans with other high-priority VA chronic conditions, such as diabetes, obesity, and depression.

B.7 Conceptual Framework

The intervention we will evaluate is based clinically on widely-adopted and evidence-based models of CBT for VA pain management (described above),[47] and it links those concepts with a strategy for personalized stepped care using **reinforcement learning (RL)**. RL is a field of artificial intelligence that allows an "intelligent agent" to learn what treatment choices work best in order to optimize a measurable outcome (termed the system's "reward"; see Figure 1). The process used to optimize treatment choices in RL mimics the way that humans learn skills such as riding a bicycle, i.e., through systematic adaptation and generalization accompanied by targeted trials of new behaviors with measurable outcomes. RL algorithms similar to those we will apply in the proposed study are the basis of online consumer targeting programs such as Netflix, Google, and Amazon.com,[48] where a service learns automatically how to deliver information that is most relevant to each user. In the current trial, the RL agent will be a computer system that makes weekly recommendations or "action choices" for each patient with respect to the mode and intensity of CBT that the patient should receive. Those recommendations will be based on the patient's progress, the progress of similar patients, and other contextual information for that action choice.

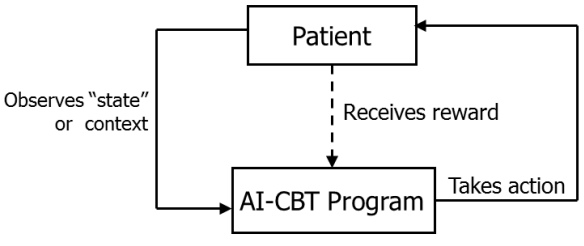

Figure 1. The Reinforcement Learning feedback loop. The AI-CBT actions are the 3 CBT session types; the "reward" is IVR-reported pedometer step-counts, and "state" data is IVR-collected information on patients' CBT skill practice & pain-related functioning.

Potential actions the AI-CBT program will recommend include: a standard one-hour telephone CBT therapy session, a 15 minute telephone CBT therapy session, and an IVR automated therapy session designed to teach and reinforce skill-based learning. Fifteen minutes was chosen to be consistent with the time increments of the health and behavior CPT codes (15, 30, 45 and 60) used to bill for behavioral interventions for chronic pain. Content for each session type will be based on standard VA CBT for pain management, modified by a Panel of Experts to be most effective given the length and mode of each contact. The AI-CBT agent's recommendation regarding which action to take will be based on each patient's IVR-reported

pedometer step counts (i.e., the "reward") as well as other "state" information (Figure 1) also collected via IVR. Importantly, the RL algorithm will learn not only based on each patient's own treatment response, but will incorporate experience from the response of other patients who have similar characteristics and response trajectories as indicated by the "state space." See Appendix A for a more detailed discussion of RL and the use of state data. Based on this feedback loop, the RL engine will modify the probability distribution across treatment choices and make recommendations for each patient each week. Because actions will be probabilistic rather than "hard-wired," the AI-CBT program will avoid potentially over-reactive treatment changes that can result when therapists attempt to tailor care non-systematically or using deterministic flow diagrams. All patients will begin with a standard one-hour CBT session. Based on their progress as measured by feedback on the "reward" and "state" features, patients who progress toward functional goals will be moved through less resource intensive options, and patients who need more intensive follow-up will be moved automatically to more time-intensive, therapist-delivered CBT.

B.8 Prior Work by the Investigators

Patient Engagement in IVR Self-Care Support Calls. Dr. Piette and his team have more than 15 years of VA experience developing IVR systems to enhance care for chronically-ill Veterans, and over the past five years, more than 2000 patients have participated in their IVR programs. Evaluations consistently indicate that these IVR-supported programs are welcomed by patients and their clinicians, and that they can improve self-care behaviors and health outcomes. In analysis of data from more than 1200 program participants with 29,000 patient-weeks of follow-up, patients completed 83% of weekly IVR assessment and self-management support calls, and completion rates were similar across groups defined by sociodemographic risk factors.[49] Other recent studies[50,51] found that: completion rates of IVR assessments are high among patients with depression, and patients' IVR reports are at least as reliable as mental health information collected via other methods. Dr. Heapy and colleagues also have found high levels of adherence to IVR call schedules in two studies among Veterans with chronic pain. In one study, they used IVR to collect daily data about pain and pain-related behaviors in a randomized clinical trial of transdermal fentanyl compared to oral short-acting opioids. Participants with endpoint data completed 85% of the expected IVR calls, and participants who withdrew or were disqualified completed 74% of potential calls.[52] In an HSR&D-funded trial, Dr. Heapy and colleagues examined the effectiveness of tailored CBT treatment for chronic low back pain and used IVR to measure patients' practice of self-management skills. Participants completed 78% of expected calls (1,856 calls completed out of 2,394 expected calls). Finally, in a pilot study designed to obtain feedback from Veterans with chronic pain about self-management, 65% indicated a willingness to receive self-management support via IVR, and 11% indicated they might be willing. The most common reason for preferring IVR support was that it would circumvent travel difficulties.

Impact of IVR Self-Management Support on the Outcomes of Chronic Illness Care. In three randomized trials directed by Dr. Piette, results indicated that IVR call-supported chronic-illness care can improve patients' self-care and health outcomes.[53-55] In a trial conducted among non-VA diabetes patients, intervention patients receiving weekly IVR monitoring and self-care support with follow-up by a telephone nurse therapist reported significantly better home glucose monitoring, foot care, medication adherence, and weight monitoring than control patients at their 12-month follow- up. More than twice as many intervention patients had acceptable glycemic control at 12 months ($p = .01$), as well as fewer diabetic symptoms, greater satisfied with care, fewer symptoms of depression, greater perceived access to care, and greater self-efficacy in managing their self-care (all $p < .05$). Veterans receiving IVR-supported telephone care management reported better self-management behaviors, were more likely to be seen in diabetes-related specialty clinics, had better glycemic control, and reported better patient-centered outcomes.

VA Industry Innovations Award. In October 2011, Drs. Heapy, Kerns, Piette and others began work on a VA Industry Innovations Award (VAi2) project in which Interactive Performance Technologies, is installing an integrated information technology infrastructure called the ePROHealth® platform on a secure server inside the VA firewall to enable IVR and text messaging-based patient self-reporting of pain relevant outcomes. The goal is for the ePRO system to have the ability to schedule regular outbound IVR pain management calls to patients, gather patient reported outcomes during those IVR interactions, store the patient reports for review by a nurse care manager, and automatically generate a CPRS note summarizing

the patient's report. This platform represents one path to dissemination if AI-CBT is found to be successful. Technical details about the platform's informatics infrastructure are included in the Resources section.

Pedometers for Monitoring Patients' Physical Activity. Dr. Piette was the Principal Investigator for an NIH-funded trial of telephone CBT plus physical activity promotion among VA and non-VA patients with diabetes and depression.[56] Investigators used pedometers to objectively measure patients' physical activity at baseline and the 12-month follow-up, and intervention patients also used pedometers as part of their CBT self-management program. We observed high rates of adherence to the collection of pedometer data, and the opportunity to use a pedometer to pursue physical activity goals was an important motivator for trial participation. Importantly, we found that intervention effects varied according to patients' disease complexity at baseline,[57] suggesting that a more personalized approach to CBT program design would optimize treatment response while using clinician time most efficiently. Caroline Richardson and colleagues have completed eight randomized trials employing pedometers and targeting VA and non-VA patients with chronic diseases (see biosketch for citations). In each study, adherence to pedometer use and reporting was high. In one ongoing trial among Veterans with COPD, 87% of intervention participants have uploaded pedometer data in the past 30 days. Drs. Sarah Krein and Caroline Richardson recently completed an HSR&D-funded trial,[58] in which Veterans with chronic back pain were randomized to a pedometer-based, Internet-mediated walking intervention or usual care. Intervention participants reported a greater decrease in back pain-related disability in the six months following study enrollment. Intervention participants uploaded pedometer data at least once per week for a median of 32 weeks (62% of the recommended time) and more than 25% of participants uploaded data at least 42 weeks (80% compliance). In summary, we have found consistently that pedometers represent an important alternative to self-reported activity levels which often have high rates of random reporting error as well as social desirability biases.[58,59]

Reinforcement Learning for Use with Mobile Health Self-Management Support. Dr. Piette and colleagues at the University of Michigan recently received two competitive grants to develop and pilot test interventions that use RL to improve the effectiveness of antihypertensive medication adherence messages delivered via text messaging. In those interventions, the "reward" (see Figure 1) used by the AI engine to make choices among message types is the patient's medication adherence as measured objectively via daily pill-bottle openings using electronic Medication Event Monitoring System (MEMS) caps. "State" data used by the AI engine in selecting among adherence support messages will include information about patients' history of systolic blood pressure readings and patients' reasons for medication non-adherence reported at baseline. "Actions" that will be automatically personalized based on patients' treatment response will be text messages addressing distinct reasons for non-adherence, including forgetfulness, concerns about side-effects, and doubts about the medication's importance.

B.9 Preparatory Collaborative Work

We performed numerous simulations in order to estimate the impact of AI-CBT compared to 10 standard, one hour CBT sessions delivered by a therapist (see Appendix A). We focused on AI-CBT's impacts on patients' physical functioning (in this case pedometer-measured step counts) and on therapist time. We assumed that (as in the proposed study) the AI-CBT program would start each patient with a one-hour therapist session and then would automatically develop a personalized step-care program that included: additional one hour "live" telephone therapy sessions, 15 minute live sessions, or IVR sessions. We assumed that patients responding to IVR therapy would also respond to a 15 minute live call or an hour long call; and patients who respond to a 15 minute therapist call (but not IVR) would also respond to an hour long session. Simulations evaluated variations in: the expected benefit of CBT delivered via different modes; the speed in which patients were recruited into the AI-CBT program (which would affect the system's ability to learn from prior experience); and whether the AI engine could move a patient directly from an hour session to IVR or whether choices were constrained so that the system would "step down" from one hour to 15 minute and from 15 minute to IVR. We also explored the effect of random error in the expected effect of each CBT session, and the effect of patients' non-adherence to the IVR requests for daily step count data (patients with more missing step count data would be progressed more slowly to less resource-intensive options). In brief, using numerous pessimistic assumptions, we estimate that AI-CBT will be able to achieve an improvement in physical activity that is more than 90% as great as that seen in

standard CBT but using less than half of the clinician time required for 10 one hour sessions for all patients.

C. SIGNIFICANCE

The proposed team includes VA researchers with a long-history of Veteran-centered health services research on mobile health, and non-VA experts with unique expertise in AI and tailored messaging that can catalyze the development of novel tools for improving chronic pain management. The use of AI to tailor patients' treatment plan is unprecedented in chronic disease management. These approaches have become widely used in other areas (e.g., online advertising) because the powerful AI algorithms can significantly improve the user-centered experience and services' impact on target outcomes. Because this approach can optimize each patient's treatment response with the most efficient use of scarce therapist time, it could dramatically increase the number of Veterans with access to pain CBT given constraints on VHA budgets. Success with this system will provide a foundation for a broader program of research on AI-supported interventions to improve chronic illness care for Veterans with high priority chronic conditions such as depression, diabetes, and obesity.

D. RESEARCH DESIGN AND METHODS

D.1 Overview

This will be a randomized non-inferiority study comparing standard pain CBT to a strategy that uses mobile health technology and artificial intelligence in conjunction with trained CBT therapists to deliver evidence-based, stepped pain therapy. Patients in both groups will receive CBT delivered via telephone by trained pain CBT therapists. For patients in the standard CBT group, the therapist will deliver 10 hour-long CBT sessions based on content used throughout VA. Patients randomized to the AI-CBT treatment group will begin with an IVR session and will be asked to report their pedometer-measured step-counts, pain-related functioning, and CBT skill practice via five-minute daily IVR calls. Some of those IVR calls also will include reminders regarding the dates and modalities for upcoming CBT sessions. Based on patients' IVR feedback, AI-CBT will make recommendations to carefully step-up the intensity of each patient's CBT follow-up using more brief telephone therapy sessions (15 minutes), or hour-long therapy sessions with the CBT therapist, if necessary. Based on experience gained from each patient's own history and the overall population of patients, the AI-CBT engine will seek to optimize the population's total improvement in functioning while maintaining each patient at the least resource-intensive mode of CBT delivery. Outcomes will be measured via telephone survey at three and six months post recruitment, and additional data will be collected via clinical records. We will use data from therapists' activity logs and administrative files to conduct a budget impact analysis. Additional data to aid translation of study findings into practice will be collected via qualitative interviews with CBT therapists, PACT team members, and patients with various levels of program response.

D.2 Patient Identification and Recruitment

Patients with diagnoses of chronic back pain receiving care in facilities affiliated with the VA Ann Arbor Healthcare System and the VA Connecticut Healthcare System will be identified via DART and CPRS. Eligible patients will have back-pain-related diagnoses including back and spine conditions, and nerve compression (using ICD-9 and -10 codes corresponding to appropriate conditions) and a score of ≥ 4 (indicating moderate pain) on the 0-10 Numerical Rating Scale on at least two separate outpatient encounters in the past year. Additionally, the pain clinic at the Ann Arbor VA Hospital will be able to inform potentially eligible participants about the study via flyer. If interested, these patients will be able to call the study and then be screened for eligibility.

Screening of interested and potentially eligible patients will be conducted by medical record review and telephone interview using validated measures. Inclusion criteria include: (1) at least moderate pain-related disability as determined by a score of 5+ on the Roland Morris Disability Questionnaire at baseline; (2) at least moderate musculoskeletal pain as indicated by a pain score of ≥ 4 on the Numeric Rating Scale; (3) pain on at least half of the days of the prior six months as reported on the Chronic Graded Pain Scale; and (4) a touch-tone cell or land line phone. Exclusion criteria include: (1) active psychotic symptoms, suicidality, severe depressive symptoms (Beck Depression Inventory (BDI) score of 30+, or a PHQ-9 score of >20),

substance use disorder or dependence, active manic episode or poorly controlled bipolar disorder as identified by Mini International Neuropsychiatric Interview,[61] or severe depression identified by chart review of diagnoses and mental health treatment notes; (2) life threatening conditions that could impede participation, such as COPD requiring oxygen or cancer requiring chemotherapy; (3) cognitive impairment defined by a score of ≤5 on the Six-Item screener;[62] (4) sensory deficits that would impair participation in telephone calls; (5) current CBT or surgical treatment related to back pain, or anything else that indicates the patient will not be able to meaningfully engage in treatment as determined by the therapist or PI; and (6) exclude if patient not planning to receive most of their pain care at the VA in the next year. After obtaining agreement from patients' primary care providers, a letter will be sent to the Veterans informing them about the study and inviting participation. Veterans who do not opt-out by postage-paid response card will be called by research staff to explain the study, conduct screening, and solicit their involvement. VA and other investigators have found that this opt-out procedure is less burdensome to patients[63] and results in a larger sample that is more representative of vulnerable patients[64] than opt-in procedures (see Human Subjects). Willing Veterans will be mailed the consent form along with a postage page return envelope. We have used this same process in numerous prior IRB-approved VA studies and have found that it is an efficient and effective way to recruit large samples of Veterans without requiring an in-person recruitment visit. In some cases, study staff may meet a Veteran in-person at AAVA to either complete the informed consent or collect their signed form rather than returning through mail. We will not require or request that Veterans come to AAVA to complete their consent form, but rather will present it as an option to them. The study coordinator will track the percentage of eligible Veterans who enroll and will actively solicit reasons for declining. This information will be used to assess the intervention's reach as described in the implementation portion of the application (Aim 3).

In preparation for this study, we used Corporate Data Warehouse records for 2012 to identify patients treated in Ann Arbor and West Haven with low back pain (ICD-9 codes 724.01, 724.02, 724.03, 724.09, 724.1, 724.2, 724.3, 724.4 and 724.5) and a pain score of ≥ 4 on the 0-10 Numerical Rating Scale on at least two separate outpatient encounters. We identified 105,344 patients and estimate that we would have to recruited 3-4% at a rate of 4.4 patients per site per month to reach accrual goals. Our prior studies based on similar populations have recruited 5 to 10 Veterans with chronic pain per month per FTE recruiter. Therefore we expect no difficulty recruiting the target sample within the proposed timeframe and staffing.

D.3 Randomization

After completing baseline assessments, patients will be randomized to AI-CBT or standard telephone CBT by the research staff. The randomization sequence and opaque randomization envelopes for recruiters will be generated prior to recruitment by the study statistician using a random number generator. Analysts and the investigators will be blinded to patients' group assignment until initial outcome analyses are completed. To ensure balance across treatment arms in potential modifiers of intervention effect, randomization will be conducted within blocks defined by site (Michigan versus Connecticut) and patient age (<65 vs. >=65).

D.4 Common Elements of Standard and AI-CBT

Both CBT conditions will consist of 10 treatment modules delivered over 10 weeks. The same therapist in each site will provide treatment to patients in both groups. In each arm, the 10-week course of therapy will include an introductory module, followed by eight pain coping skills training modules and concluding with a session emphasizing skill consolidation and relapse prevention. The introductory module will present the biopsychosocial model which explains how chronic pain leads to dysfunction and provides a rationale for the efficacy of pain coping skills. The eight skills that will be presented were selected based on their efficacy in improving pain outcomes and their appeal to patients in prior trials. These include sessions on: physical activity, behavioral activation, pacing, sleep hygiene, and relaxation. Other modules will address common maladaptive cognitions including pain catastrophizing and fear of movement or kinesiophobia. Using procedures developed in two previous VA studies, during sessions 2-9 participants will be assigned a goal related to a newly presented skill (e.g., "practice relaxation exercise for 20 minutes daily") and a daily walking goal (e.g., average daily steps over the prior week plus 10%). As participants progress through treatment, they will continue to practice prior goals. In order to maintain consistency in the behavioral and

cognitive restructuring targets of CBT between the two treatments, participants in both groups will be assigned the same skill practice goals and the same formula will be used for assigning steps goals.

Patient and Therapist Materials. Patients in both treatment conditions will use a handbook that was developed from materials used in prior trials, refined for use with IVR CBT through an HSR&D-funding Short Term Project Award, and currently in use in our HSR&D-funded IIR of IVR-based CBT for low back pain (see appendices). The handbook is written at the 6th grade reading level so that it can be used with or without the guidance of a therapist. The handbook will be identical for both conditions except that the AI-CBT handbook will contain additional information that describes the three AI modes (one hour, 15 minute, and IVR sessions) and how to prepare for each type of session. The therapist manuals will be adapted from materials developed for our IVR-based CBT for Chronic Low Back Pain trial (see appendices). The AI-CBT section will detail specific guidelines for each treatment mode: one hour, 15 minute, and IVR.

Pedometers for Monitoring Physical Activity. All patients will be given a pedometer and a log for monitoring their step counts. We will use a Yamax DigiWalker pedometer because it is accurate and used frequently in research.[65] Patients will be mailed a pedometer after completing their baseline assessment and returning their consent form.

D.5 Standard Telephone CBT (control)

Control patients will receive telephone CBT consisting of 10 weekly modules delivered via one hour telephone contacts with a therapist. The format of each session will include: (1) review of patients' pedometer logs and coping skill practice, (2) review of previous material and correction of misunderstandings of the information, (3) assignment of new step count goals and discussion of the new skills-based material, and (4) discussion of specific step and skill practice goals. Positive feedback and praise will be offered for any skill practice and step goal efforts and accomplishment. Barriers to practice or goal completion will be identified and problem-solving techniques will be used to address them.

D.6 Cognitive Behavioral Therapy Supported by Artificial Intelligence (AI-CBT)

Daily IVR Reports. We will collect reports about patients' pedometer-measured step counts, CBT skill practice, and pain-related functioning via daily five-minute IVR calls. If the initial call is missed, the system will automatically try again 15 minutes later and again 1 hour later. We have successfully used these methods in studies achieving high patient response rates. Step counts will measure activity over the prior 24 hours, and patients will report their skill practice using a 0 -10 scale. Pain-related functioning will be assessed using a single item from the West Haven-Yale Multidimensional Pain Inventory (WHYMPI).[66] Practice of target behaviors will be measured using a 0 (not at all accomplished) to 10 (completely accomplished) scale and items that we have successfully used in our ongoing trial if IVR CBT. Step counts will be used in the "reward function" that the RL algorithm will seek to optimize; and functioning and skill practice reports will be used as "state" information (see Appendix A) that the system will take into account when making decisions that optimize patients' treatment course. The AI system will be able to accommodate missing IVR reports, and patients who fail to complete more than 50% of the daily IVR calls in a two week period will be called by a research associate to trouble shoot problems and encourage compliance with feedback. In addition to being the source of data with which AI-CBT will personalize each patient's course of treatment, web-based reports of data from IVR calls will be used to inform therapists of participants' progress. These data will be particularly important for informing the abbreviated 15 minute therapist sessions where a priority will be placed on the efficient use of treatment time, and during the IVR sessions where the entire session will be pre-recorded. Once a week the IVR call will include a brief message alerting patients of the date, time, and modality for their subsequent week's session.

AI-CBT Action Recommendations. After week 1, session options will include: (a) one-hour telephone treatment sessions; (b) 15 minute live telephone therapist sessions; and (c) IVR treatment sessions. To avoid scheduling conflicts, AI-CBT patients will be assigned a one-hour block of time each week in which both they and the CBT therapist are available for treatment. This same time slot will be used for either the hour-long therapist sessions, the 15 minute therapist sessions or the IVR CBT sessions. Each Monday morning, the CBT therapists will receive a list of AI-CBT personalized treatment recommendations for that

week for each patient. By noon on Monday, the therapist will have a finalized schedule of which patients require what types of contact that week, and which patients need to have a summary of the therapist's comments and recommendations recorded for the week's IVR CBT therapy call.

During *Week 1*, AI-CBT patients will have an IVR session where they will be introduced to the program's goals and processes, and the standard introductory material contained in session 1 of the standard CBT via the IVR system. The *One Hour AI-CBT Sessions* will be identical to the control condition and will follow the same progression of content used for control patients. *15 Minute Telephone CBT Sessions* will mirror the content of the one-hour sessions, though in a compressed form. The therapist manual and patient handbook will emphasize the importance of using the session time efficiently and using a consistent format that includes: reviewing the patient's daily IVR reports, clarifying the current week's adaptive pain coping skill in the patient handbook, and setting goals for skill practice and step counts for the coming week. Prior to the session, therapists will review patients' daily IVR reports. If participants have not been successful in meeting step or skill practice goals, the therapist will help the participant address barriers to goal attainment. The therapist then will ask the patient to describe the current week's adaptive pain coping skill as a brief check of their understanding, and will review goals for the coming week with a discussion of anticipated barriers. Remaining time will be used to review the skill and to encourage the patient to read their patient handbook. Much of the content for the *IVR CBT sessions* has been developed and implemented as part of our ongoing IVR-based CBT for Chronic Low Back Pain trial (see Appendix D). Our experience in that trial suggests that patients complete the IVR sessions more than 90% of the time and that satisfaction rates are high. During these sessions, patient will receive a 2-5 minute pre-recorded feedback message from their therapist, during which the therapist will review the patient's IVR-reported changes in step counts, functioning, and skill practice. Reinforcement will be provided for effort, and improvements will be noted. IVR messages will include a review of the pain coping skill practice and step goals for the coming week, and participants will have the option of leaving a message for their therapist via the IVR system should they have a question. Therapists can leave a response message, also on the IVR system.

The AI Engine. Patients' IVR-reported step counts, skill practice, and pain-related functioning will be accessed by the AI engine daily to update the probabilities the system uses to determine which treatment to recommend the next week. We will use a state-of-the-art "contextual bandit" AI algorithm (LinUCB) designed to make careful choices while learning quickly from a patient's treatment response as well as the experience of other patients with similar characteristics (see Appendix A for technical details).[67-70] With increased interactions, the system will learn to tailor decisions more effectively to maximize population-level improvements in functioning while minimizing clinician time. In this way, AI-CBT will function similar to the best clinicians, who learn from experience within and across patients. In the context of the trial, this means that patients enrolled early will likely receive less personalized CBT courses that are relatively similar to the standard CBT approach (i.e., a greater number of hour-long sessions), while patients enrolled later will receive services that are more personalized and include a greater frequency of 15 minute therapist sessions and IVR sessions. To maximize the efficiency of this "learning curve": (1) patients will be recruited over a longer period than would potentially be necessary, so that AI-CBT can gain as much experience as possible from patients recruited first and apply that knowledge to patients entering the program later; and (2) patients will be randomized with a greater "N" in the AI-CBT group in order to maximize the system's experience (see power calculation). As part of our evaluation, we will compare outcomes across randomization groups separately for early versus later enrollees, in order to test the hypothesis that AI-CBT will result in greater efficiency over time. These analyses also will allow us to project program benefits if AI-CBT were implemented in VA for thousands of Veterans and multiple years of experience.

CBT Treatment Fidelity will be assessed using the Yale Adherence and Competence Scale (YACS),[71] a validated scale that assesses therapist adherence and competence in delivering manualized behavioral therapy. Dr. Heapy will rate audiotapes of 30% of all CBT therapist sessions to assure that treatment is consistent with the manual and will provide corrective feedback to the therapists whenever drift occurs.

Role of the Expert Panel. The AI-CBT program will be supervised with ongoing input from an Expert Panel comprised of experts in: pain management, CBT for chronic pain, clinical trials using behavioral interventions, adaptation of psychotherapy for telephone delivery, and IVR (see Appendix E and letters of

support). The Panel will be directed by Dr. Robert Kerns (Co-I) who has led many large and geographically-dispersed panels of experts in pain research and treatment, and will ensure that the panel meetings are efficient and effective. The Panel will meet regularly by teleconference during the start-up period to review and revise the treatment materials and refine the AI algorithm to reflect any constraints that should be put into place to limit the choices the algorithm can make, e.g., "if the patient's activity level decreases more than 20% two weeks in a row, recommend two hour-long CBT session regardless of what the prior week's contact was." After start-up, the Panel will teleconference quarterly and in ad hoc sessions as needed.

D.7 Measurement

Overview. We have selected outcome measures based on recommendations from the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations.[72,73] Endpoint measures are consistent with CONSORT guidelines recommending that non-inferiority trials use outcomes similar to those used in efficacy studies. In addition to primary and secondary outcomes (defined below), we will examine treatment satisfaction, treatment credibility, patient engagement and dropout, and goal accomplishment. Process and outcome data will be collected via the following sources: *Patient Surveys:* Baseline, three-month, and six-month surveys will be conducted via telephone by trained research assistants. *Qualitative Interviews:* These will be conducted with purposive samples of patients in the AI-CBT group at follow-up. We will target patients who demonstrate significant improvement, patients who were very satisfied with AI-CBT, patients without significant improvement, patients who were dissatisfied, and patients who dropped out of the intervention. Additional qualitative interviews will be conducted at follow-up with CBT therapists and PACT team members. *CBT Therapist Logs:* Logs will be used to track therapist time spent in patient treatment, attempting to reach patients, and key information about those interactions. *IVR:* The AI-CBT IVR system will capture information about intervention patients' pedometer-measured step-counts, pain related functioning, CBT skill practice, and missed data reporting. *Administrative and Clinical Data Systems:* These will be used to track patients' use of other VA inpatient and outpatient services for pain management, mental health, and medical care.

Primary Outcome. The 24-item Roland Morris Disability Questionnaire (RMDQ) is an IMMPACT endorsed measure[73] of pain-related disability for persons with chronic low back pain. Strong evidence supports the RMDQ's reliability, validity, and responsiveness to change during trials.[74]

Secondary Outcomes. *Global Pain Intensity* will be assessed using the Numeric Rating Scale (NRS-I) an IMMPACT-recommended 11-point numeric rating scale of pain severity.[72] *Pain-Related Interference* will be measured using the 9-item Interference subscale of the West Haven-Yale Multidimensional Pain Inventory (WHYMPI).[66] This IMMPACT-recommended measure assesses pain-related interference in daily activities and has demonstrated good internal consistency.[72] Depression Symptom Severity will be assessed using the 21-item Beck Depression Inventory (BDI), a widely used measure with excellent internal consistency and stability.[60] *The Patient Global Perception of Change* scale is a single item measure that quantifies a participant's overall perception of improvement since beginning treatment and the clinical importance of that improvement. Participants indicate improvement on a 7 point "much worse" to "much better" scale. This is a well-validated measure recommended by IMMPACT.[73] Finally, we will use the *Veterans SF-12 to assess health-related quality of life*. This measure has demonstrated good internal consistency and is strongly correlated with socioeconomic status and morbidities.[75]

Resource Use (Aim 2). *Intervention Costs.* Therapists will use a log to record time spent in intervention-related activities for a random 20% of all treatment days. Time records will be combined with wage data from the VA Financial Management System to estimate intervention-specific personnel costs. Technology costs of the AI-CBT program include fixed costs (e.g., software development and computer maintenance) plus variable costs (e.g., minute costs for IVR calls). One-time fixed start-up costs will be reported separately. *VA Inpatient and Outpatient Service Use* data will be obtained from the Musculoskeletal Diagnoses Cohort (MSD), a project currently underway as part of the VA Connecticut Healthcare System's CREATE. The MSD is developing validated algorithms for using VA electronic health record data to identify utilization events, comorbid conditions, receipt of opioid medications, and pain screening results, for patients with pain-related diagnoses. Information on non-VA admissions will be collected by the patient

survey. To mitigate recall bias, we will use a two-time frame method that asks about utilization over the past 6 months and over the past 2 months with more weight given to the shorter timeframe.[76]

Treatment Satisfaction and Engagement (Aim 3). For patients in the AI-CBT group, we will calculate *IVR adherence* as we have in the past,[49] i.e., as the proportion of days during which an assessment was attempted in which one was successfully completed and the number of weeks during which the patient completed at least four out of seven requested IVR reports. Participants' judgments of *Treatment Credibility* will be assessed using a reliable questionnaire adapted from Borkovec and Nau.[77] Treatment credibility has been shown to be significantly associated with treatment satisfaction, engagement in treatment, and number of sessions attended. The *Pain Treatment Satisfaction Scale* of the Patient Outcomes Questionnaire will be used to assess patient satisfaction with various domains of pain care.[78] This 5-item measure shows good internal consistency and significant associations with staff and patient ratings of patient improvement. *Attendance in "Live" Telephone CBT Sessions and Program Dropout* In order to understand reasons for treatment dropout, we will attempt to reach samples of patients with low levels of engagement for qualitative interviews. Participants will rate their *Continued Skill Use* at follow-up for each of the target behaviors emphasized in the CBT program on a 0 (not at all accomplished) to 10 (completely accomplished) scale. As described above, *Daily IVR calls* will be used to collect data in the AI-CBT condition regarding pedometer measured step-counts, CBT skill practice, and pain-related functioning using pre-recorded questions we have used successfully in our prior studies.

Demographics and Other Covariates Measured at Baseline. We will measure patients' baseline *Sociodemographic and Pain Characteristics* that have been shown to be associated with treatment outcomes, e.g., age, gender, education level, racial/ethnic background, marital status, occupational status, pain duration, and number and location of pain sites. We also will gather data on participants' level of health literacy.[79] *Pain Classification* will be derived through a systematic evaluation of each enrolled participant's EMR using an assessment tool based on clinical guidelines for diagnosing and treating back pain.[1] Pain classes will include non-specific back pain, back pain with a radicular component, or back pain associated with other specific spinal causes. A nurse practictioner, who is supported by the West Haven COIN and who has been trained to use this tool and is using it in our trial IIR 09-058, will conduct the assessment. She will be supervised by a pain medicine physician (Dr. Rosenberg) who will review ten percent of the classification ratings for accuracy and provide corrective feedback if necessary. *Psychiatric and Substance Abuse Comorbidities* will be measured using medical record diagnoses and mental health encounters. Additional self-report information will be collected using subscales of the Mini International Neuropsychiatric Interview (MINI)[61] related to mood and substance abuse disorders. *Pain Medication Use* will be assessed through patient survey and a review of computerized pharmacy records. Pain medication will be coded as non-steroidal anti-inflammatory, non-narcotic analgesics, narcotic analgesics, and benzodiazepines and other sedative/hypnotics. For each category, we will document post-treatment whether patients have experienced an increase, no change, or decrease in their medication use. *Distance from VA* will be calculated and used as a measure of geographic access. The *Pain Catastrophizing Scale* is a 13-item self-report scale that examines thoughts and feelings people may experience when they are in pain including rumination, magnification, and helplessness.[80] Finally, *Pain-related Fear* will be measured using the Tampa Scale of Kinesiophobia-revised (TSK-R), which has two subscales (Fear of Harm/Activity Avoidance and Pathophysiological Beliefs) and has been shown to be sensitive to treatment-related change.

D.8 Sample Size and Power Calculation

Our primary outcome analyses are consistent with the CONSORT Statement on Reporting of Noninferiority and Equivalence Randomized Trials.[81] The sample size was calculated using the Non-Inferiority Test module available in the statistical software PASS 2008. The calculation was reviewed by the study statistician, Dr. Myra Kim. To ensure that the AI-CBT program retains a clinically acceptable effect, the non-inferiority margin was set at 2 points on the Roland Morris Disability Questionnaire (RMDQ).[74] A two point difference (or difference in reduction) in the RMDQ is considered to be a minimally clinically significant effect.[82] The power calculation was based on a significance level of 0.025, with a power of 90%, when postulating a true difference in group means of 0 and a standard deviation of the outcome of 4.5 in both groups. Specifically, if we denote by D the true difference in mean RMDQ scores (at 12 weeks) between the

AI-CBT and standard CBT groups, with the non-inferiority margin set at 2, we plan on testing the null hypothesis H0: D>=2 versus the alternative hypothesis H1: D<2, which amounts to a one-sided, two-sample t-test. Thus, if the null hypothesis is rejected, it can be concluded that AI-CBT is non-inferior to CBT. We will test this hypothesis at a one-sided .025 significance level based on a confidence level where we will reject the null hypothesis when the upper bound of the two-sided 95% confidence interval for D is less than 2. To ensure that the AI-CBT algorithm has as much information as possible to learn quickly how best to tailor patients' course of therapy, we will disproportionately randomize patients to the AI-CBT group in an allocation ratio of 1.37:1 for the AI-CBT: standard CBT groups, respectively. Assuming this ratio, we will have 90% power to detect non-inferiority with a total sample size of 221 patients, or 128:93. To account for a 20% drop-out rate in both groups, we will enroll 278 patients (160 in the AI-CBT group, 118 in CBT group).

D.9 Analysis

Baseline Comparability, Reach, and Representativeness. We will examine baseline differences across groups in measures of study endpoints as well as other potential prognostic indicators, such as patients' age, comorbid diagnoses, and history of pain treatment. Any differences across groups in baseline characteristics will be controlled statistically in analyses comparing outcomes. The RE-AIM framework is a methodology for systematically considering all strengths and weaknesses of an intervention in order to better guide program planning.[83] To evaluate reach, we will ask patients who decline study participation whether they would be willing to provide informed consent to participate in a brief survey that identifies their reasons for declining participation and the characteristics that differentiate them from enrollees.

Analysis of Endpoints (addressing Specific Aim 1). We will test for non-inferiority of AI-CBT compared to standard telephone CBT at 12 weeks by comparing the upper limit of the two-sided 95% confidence interval for the difference in the mean RMDQ scores, calculated as AI-CBT minus standard CBT, to the pre-set non-inferiority margin of 2 points in RMDQ and will conclude that AI-CBT is non-inferior if the upper limit is less than 2. Because intent-to-treat analysis can raise the risk of type I error in a non-inferiority trial,[81] we will conduct both a per protocol and intent-to-treat analysis. The "per protocol" group assignment will be defined as completing four or more CBT sessions (either IVR or "live" sessions); however, we will revisit this definition with the Expert Panel prior to the start of the trial. We will declare AI-CBT non-inferior to standard CBT only if AI-CBT is shown to be non-inferior using both the intention-to-treat and per-protocol analysis sets. We expect that RMDQ scores will be normally distributed. If not, we will use transformations to achieve normality. We also will develop a two-level linear mixed-effects model that uses RMDQ follow-up scores at both assessment times (three and six months) as the dependent variables; treatment group, time and the treatment by time interaction as categorical explanatory variables; and baseline RMDQ score as a continuous covariate. If this model shows no significant time by group effect, we can drop the interaction term and test for the time-averaged non-inferiority of AI-CBT compared to standard CBT between 3 and 6 months. This model will also allow for adjustment for design-related factors (e.g., site and age). An unstructured variance-covariance matrix will be used to model the error variance. Secondary outcomes including pain intensity, emotional functioning, global perception of change, and quality of life will be analyzed in a manner similar to that used for the primary outcome. Analyses of all other outcomes will be conducted on an intent-to-treat basis.

Intensity of Service Use (Specific Aim 2). We will compare service utilization by category (e.g., CBT therapist time, PCP visits, and pharmacy use) between groups. We will conduct a budget impact analysis[84] and will include the cost of the intervention (personnel, supplies, CBT therapist training, and IVR fixed/variable costs) as well as costs for specific medical care services likely to be affected. Data from CBT therapists time records will be combined with wage data from the VA Financial Management System to produce estimates of intervention-specific personnel costs. Costs associated with the use of specific medical care services, such as medications, will be obtained from the Decision Support System (DSS) files. Cost analysis will be conducted in accordance with the guidance provided by Mauskopf et al.[84] including the use of sensitivity analysis and scenarios that allow for varying assumptions about intervention uptake, compliance or component costs. All resource use and cost comparisons will be adjusted for any

observed differences in baseline characteristics. Because costs of resource utilization are usually skewed, alternative modeling techniques (e.g., log-transformed costs, negative binomial regression) will be used.

Intervention Engagement and Satisfaction with Care (Specific Aim 3). As in our prior research,[49] we will conduct extensive analyses of the process of intervention delivery in both arms. We will monitor the proportion of telephone CBT sessions that are completed, and we will determine the patient and session characteristics associated with patients' reports of skill practice. Patients in the AI-CBT group will report their satisfaction with aspects of the intervention (e.g., whether it provided information useful for achieving behavioral targets), and we will assess the correlation between satisfaction ratings and measures of: intervention engagement, patients' baseline characteristics, and changes in pain-related functioning. Differential dropout across groups will be examined using Kaplan-Meier curves and survival models.

Preplanned Subgroup Analysis. Because AI-CBT will continue to learn patterns in patients' experience throughout the intervention period, we hypothesize that the second 50% of patients randomized will show an even larger difference in clinician time than the first 50%, while still demonstrating non-inferiority in pain-related outcomes. Differences in pain related functioning and in clinician treatment time across treatment groups will be tested in this subgroup analysis after stratifying the sample into early versus later recruits.

Approach to Missing Data. If more than 15% of a covariate is missing, we will use multiple imputation methods based on the SAS MI Procedure. We will check if the pattern of missingness is monotonic (i.e., if patients missing data at 3 months also have missing data at 6 months) and use a Markov chain Monte Carlo method that assumes multivariate normality to impute missing values. We will impute five sets of data and combine the imputed results using the SAS MIANALYZE procedure to obtain a valid estimate of the confidence limit and treatment effect.[85] When data are missing for items within scale scores, we will use recommended imputation procedures rather than deleting patients list-wise from the analysis.

Mediators and Moderators of Intervention Effects. We will use multivariate modeling to identify the mechanisms through which the intervention achieves effects on outcomes and differential effects across subgroups.[86] Initial models will include only treatment group as the predictor. Subsequent nested models will introduce potential mediators (such as the number of completed therapist sessions), and we will evaluate changes in the relationship between experimental condition and outcomes before and after covariates are introduced. Analyses of effect moderation will focus on baseline pain severity and comorbid diagnoses using standard approaches to evaluate interactions between these covariates and patients' experimental condition.[86] Significant interactions will be interpreted by plotting regression lines for predicted outcomes of patients with high and low values of the moderator.

Evaluating the Reliability of Patients' IVR Reports. We will evaluate the integrity of IVR-reported step counts and functioning by examining associations between IVR reports and baseline characteristics that the literature suggests would be associated with patients' functioning (e.g., baseline SF-12 scores, comorbid medical diagnoses, and age). We also will examine serial correlations across IVR reports under the assumption that all correlations between scores and proximal scores should be positive and roughly of equal magnitude controlling for the time difference between reports.

RE-AIM[83] Dimensions of Intervention Adoption, Implementation, and Maintenance. Adoption will be evaluated by examining variation in study participation and intervention engagement across sociodemographic and clinical subgroups of eligible patients. For example, we will determine whether older patients or those with less education have more difficulty responding to queries about their step counts or pain-related functioning via IVR. Adoption at the provider level will be monitored by recording the proportion of providers who are willing to have their patients participate and providers' reasons for not participating. Implementation and maintenance will be evaluated through semi-structured questions at follow-up designed to identify program characteristics that might be a barrier to patients' use of the intervention in other settings and the intervention characteristics that patients feel would make it more valuable to others with chronic pain. We also will meet with clinicians in each site to gauge their willingness to adopt and maintain a similar intervention, and the ways such a system can be designed to best complement existing services.

Qualitative Interviews and Mixed-Methods. We will use audio-taped interviews with patients, CBT

therapists, and PACT team members to provide a context for interpreting intervention effects and suggest additional subgroup analyses. Interviews will be transcribed verbatim, and we will enter the transcripts into NVivo for file storage and selective retrieval. Using accepted techniques,[87] Drs. Piette and Heapy will independently read transcripts, approaching the data with analytic categories in mind, but identifying other categories in the data. An iterative process will be used until agreement is reached on categories and their definitions, after which we will develop a coding template and enter it into NVivo as a tree diagram.

D.10 Limitations of the Design

The AI engine will only make productive decisions about patients' therapy if valid and reliable information about patients' progress is available. In this respect, the AI-CBT program is identical to clinicians who must rely on patient reports to judge treatment response after a change in management. The AI engine will be programmed to treat dramatic changes in step counts that are likely to be inaccurate as missing data. Missing data on step counts and CBT skill practice will result in conservative choices in the AI-CBT group which in the extreme will leave patients with weekly one hour telephone CBT identical to that received by control patients. In contrast to other applications of reinforcement learning (e.g., purchases among Amazon.com users), the AI engine in this intervention will only be receiving data on a relatively small number of patients and time points. As such, the system will learn relatively slowly for a given patient, especially for patients enrolled in the initial phases of recruitment. We expect that this will lessen the system's ability to maximize cost-savings by offering less resource intensive but equally effective alternatives to extended telephone CBT sessions. As such, the differences in per-patient treatment cost across groups will be a lower bound of what could be expected if the service were implemented in VA with larger samples of patients over longer periods of time. Veterans with chronic pain who have significant comorbidity or are not able to walk will be excluded, limiting generalizability. Sensor technology is continuing to evolve rapidly, and the feedback from IVR+pedometers will likely under-estimate the ability of AI-CBT to personalize treatments using more passive and accurate activity measures yet-to-be-developed.

E. DISSEMINATION AND IMPLEMENTATION PLAN

The dissemination plan will be guided by the Replicating Effective Programs (REP) framework.[88] The REP framework includes guidelines within four stages for program implementation and dissemination: pre-conditions (e.g., identifying need, the target population, and a suitable intervention); implementation (e.g., package dissemination, training, and technical assistance), maintenance, and evolution (e.g., fidelity assessment). Given the state-of-the-science, we will focus mainly on gathering and disseminating information related to the implementation and potential program maintenance as a sustainable, inter-compatible program within VA's informatics environment. Specific information related to the REP framework will be identified via a dialogue with experts in VA health informatics. We have added Dr. Cynthia Brandt to the research team (see biosketch and support letter), who will be instrumental in ensuring that the AI-CBT intervention is scalable and compatible with other VA research and operational informatics initiatives. We also have reached out the eHealth QUERI (see letter of support) to guide our implementation/dissemination activities, both as a source of information about VA informatics activity nationally, as well as a conduit for sharing study results with the field. Resources from the West Haven PRIME Center, directed by Dr. Robert Kerns, will be made available to assist with dissemination via websites for PRIME and the National Pain Program Office. An additional avenue for dissemination is the National Pain Research Working Group, which includes VA and non-VA pain investigators that teleconference regularly to identify priorities for pain research and develop collaborative projects. Dr. Piette is a member of the QUERI-Diabetes program and of the Serious Mental Illness Treatment Research and Education Center (both in Ann Arbor), and directs the patient-centered technology core for the Ann Arbor COIN. In these roles, he has close ties with the national PACT leadership for exploring potential demonstration studies. The Center for Information Dissemination and Education Resources (CIDER) and the Employee Education System will be used to target VA policy makers through cyber seminars. Dr. Kerns is a longstanding member of the Professional Advisory Board for the American Chronic Pain Association (ACPA), the largest patient advocacy organization in the US, and we will use this linkage to engage consumers in implementation and dissemination efforts.

F. PROJECT MANAGEMENT AND TIMELINE

The Co-PIs will teleconference at least weekly. Meeting minutes will be kept in a shared folder on the Ann Arbor COIN VA shared drive, accessible to all investigators and Expert Panel Members. Teleconferences with Co-Is will be held monthly in Year 1 and bi-monthly thereafter, with additional meetings as needed. The Expert Panel will meet with the Co-PIs by teleconference monthly during start-up and then quarterly during the rest of the study. Ad hoc meetings of the Panel will be held if questions arise during data collection regarding the implementation of the AI-CBT intervention. Data collection manuals will describe issues such as how to note missing data and make changes on data collection forms. A coding decisions log will be maintained to ensure coding is consistent across experimental conditions and over time. All data forms will be edited by the Project Manager, double entered, and checked for out-of-range values and logical consistency across variables. Additional information about the current IVR platform and the ePRO system that we hope to use for implementation (see Section B.8) is included in the Facilities and Resources Section.

Figure 2: Project Timeline	**Year 1**				**Year 2**				**Year3**				**Year4**			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
IRB submission																
Finalize protocols and materials																
Draw patient data																
Screening and recruitment																
Intervention delivery																
6 month assessment																
Data analysis																
Report writing/dissemination																

Data Analyses Plan
IIR 13-350 Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools

Table of Contents

Data Analyses Plan

IIR 13-350 Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools

Specific Aims and Hypotheses

Chronic low back pain is one of the most common health problems among VA patients and is associated with emotional distress, substance abuse, suicide, and difficulty managing comorbid medical conditions. **Cognitive behavioral therapy (CBT)** is one of the most widely accepted, evidence-based approaches to pain management. Unfortunately, VA CBT programs are labor intensive, typically requiring 10 hour-long weekly sessions to teach self-management skills. Because of the scarcity of trained therapists and the high service costs, only half of Veterans receive care in a facility where pain CBT is offered. Where services are available, many Veterans drop out of treatment early, either because they experience minimal additional improvement after the first few sessions or because they simply cannot tolerate repeated in-person visits. Moreover, CBT quality may vary across therapists and patients due to the lack of a systematic method for monitoring Veterans' progress and personalizing treatment in light of their unique trends. Effective CBT programs vary widely in their program length and mode of delivery (in-person therapy, telephone, or even on-line), and it is likely that one size does not fit all patients' needs. For all of these reasons, the current VA delivery models are out-of-step with VA's patient-centered philosophy, and inconsistent with a stepped care approach designed to make service delivery more effective and efficient based on systematic evaluations of each patient's needs. As a result, current programs may expend scarce resources on some patients who could achieve equal benefit (or even greater benefit) from a more personalized approach that takes into account evidence regarding each Veteran's progress when planning subsequent sessions.

Mobile health (mHealth) technologies including **interactive voice response (IVR)** calls may increase access to pain-related patient monitoring and self-management support. Studies by us and others have demonstrated that: patients with pain conditions, psychiatric illnesses, and chronic medical disorders can and will use mHealth services to report valid and reliable clinical information[5,6]*;* and that mHealth services can improve self-care behaviors and physiologic outcomes. However, like current CBT pain services delivered by "live" therapists, mHealth interventions almost universally use either simplistic messages that are invariant across patients and over time, or they use pre-defined, tree-structured algorithms to determine which messages to send to which patients and how often those messages occur. Because these rigid systems cannot adapt to complex, time-varying patient needs, mHealth programs are often perceived as "robotic" by users, and patient engagement wanes over time.

The current study will evaluate a new approach for delivering stepped care for CBT pain management that may substantially improve the ability of VA pain CBT to systematically adapt to patients' unique needs while ensuring that scarce clinician resources focus on patients who need more intensive treatment. Specifically, we will use **Reinforcement Learning (RL)**, a type of **Artificial Intelligence (AI)** to ensure that Veterans with chronic pain receive the intensity and type of therapy they need, while automatically stepping patients down to less resource-intensive service options if that patient demonstrates equal or greater benefit using those approaches. This project is the result of a unique collaboration between VA experts in pain management, mHealth, and self-management support; as well as University of Michigan experts in AI and tailored health communication. Our *central hypothesis* is that a pain management program that uses AI algorithms and regular feedback collected via IVR about Veterans' physical activity, CBT skill practice, and pain-related physical functioning will automatically adapt treatment delivery to achieve outcomes that are as good as standard approaches (or even better), but using substantially less clinician time. The specific aims of the study are as follows:

(1) Using a randomized non-inferiority design, we will determine whether an **AI adaptive program of telephone-based, personalized and stepped pain CBT that includes brief therapy sessions and automated calls as well as standard one hour treatment sessions (AI-CBT)** has outcomes that are non-inferior compared to a standard CBT program of 10 hour-long sessions;

(2) We will use a budget impact analysis to quantify the difference in therapist time associated with AI-CBT relative to standard CBT as well as differences in overall ambulatory and inpatient service use;

(3) We will determine whether AI-CBT results in greater patient satisfaction and engagement, due to its more patient-centered approach that automatically focuses on "what works" for each Veteran.

Sample Size Rationale

Our primary outcome analyses are consistent with the CONSORT Statement on Reporting of Noninferiority and Equivalence Randomized Trials.[81] The sample size was calculated using the Non-Inferiority Test module available in the statistical software PASS 2008. The calculation was reviewed by the study statistician, Dr. Myra Kim. To ensure that the AI-CBT program retains a clinically acceptable effect, the non-inferiority margin was set at 2 points on the Roland Morris Disability Questionnaire (RMDQ).[74] A two point difference (or difference in reduction) in the RMDQ is considered to be a minimally clinically significant effect.[82] The power calculation was based on a significance level of 0.025, with a power of 90%, when postulating a true difference in group means of 0 and a standard deviation of the outcome of 4.5 in both groups. Specifically, if we denote by D the true difference in mean RMDQ scores (at 12 weeks) between the AI-CBT and standard CBT groups, with the non-inferiority margin set at 2, we plan on testing the null hypothesis H0: D>=2 versus the alternative hypothesis H1: D<2, which amounts to a one-sided, two-sample t-test. Thus, if the null hypothesis is rejected, it can be concluded that AI-CBT is non-inferior to CBT. We will test this hypothesis at a one-sided .025 significance level based on a confidence level where we will reject the null hypothesis when the upper bound of the two-sided 95% confidence interval for D is less than 2. To ensure that the AI-CBT algorithm has as much information as possible to learn quickly how best to tailor patients' course of therapy, we will disproportionately randomize patients to the AI-CBT group in an allocation ratio of 1.37:1 for the AI-CBT: standard CBT groups, respectively. Assuming this ratio, we will have 90% power to detect non-inferiority with a total sample size of 221 patients, or 128:93. To account for a 20% drop-out rate in both groups, we will enroll 277 patients (160 in the AI-CBT group, 117 in CBT group).

Randomization

After completing baseline assessments, patients will be randomized to AI-CBT or standard telephone CBT by the research staff. The randomization sequence and opaque randomization envelopes for recruiters will be generated prior to recruitment by the study statistician using a random number generator. Analysts and the investigators will be blinded to patients' group assignment until initial outcome analyses are completed. To ensure balance across treatment arms in potential modifiers of intervention effect, randomization will be conducted within blocks defined by site (Michigan versus Connecticut) and patient age (<65 vs. >=65).

Description of Data Collection

Overview. We have selected outcome measures based on recommendations from the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations.[72,73] Endpoint measures are consistent with CONSORT guidelines recommending that non-inferiority trials use outcomes similar to those used in efficacy studies. In addition to primary and secondary outcomes (defined below), we will examine treatment satisfaction, treatment credibility, patient engagement and dropout, and goal accomplishment. Process and outcome data will be collected via the following sources: *Patient Surveys:* Baseline, three-month, and six-month surveys will be conducted via telephone by trained research assistants. *Qualitative Interviews:* These will be conducted with purposive samples of patients in the AI-CBT group at follow-up. We will target patients who demonstrate significant improvement, patients who were very satisfied with AI-CBT, patients without significant improvement, patients who were

dissatisfied, and patients who dropped out of the intervention. Additional qualitative interviews will be conducted at follow-up with CBT therapists and PACT team members. *CBT Therapist Logs:* Logs will be used to track therapist time spent in patient treatment, attempting to reach patients, and key information about those interactions. *IVR:* The AI-CBT IVR system will capture information about intervention patients' pedometer-measured step-counts, pain related functioning, CBT skill practice, and missed data reporting. *Administrative and Clinical Data Systems:* These will be used to track patients' use of other VA inpatient and outpatient services for pain management, mental health, and medical care.

Primary Outcome. The 24-item Roland Morris Disability Questionnaire (RMDQ) is an IMMPACT endorsed measure[73] of pain-related disability for persons with chronic low back pain. Strong evidence supports the RMDQ's reliability, validity, and responsiveness to change during trials.[74]

Secondary Outcomes. *Global Pain Intensity* will be assessed using the Numeric Rating Scale (NRS-I) an IMMPACT-recommended 11-point numeric rating scale of pain severity.[72] *Pain-Related Interference* will be measured using the 9-item Interference subscale of the West Haven-Yale Multidimensional Pain Inventory (WHYMPI).[66] This IMMPACT-recommended measure assesses pain-related interference in daily activities and has demonstrated good internal consistency.[72] Depression Symptom Severity will be assessed using the 21-item Beck Depression Inventory (BDI), a widely used measure with excellent internal consistency and stability.[60] *The Patient Global Perception of Change* scale is a single item measure that quantifies a participant's overall perception of improvement since beginning treatment and the clinical importance of that improvement. Participants indicate improvement on a 7 point "much worse" to "much better" scale. This is a well-validated measure recommended by IMMPACT.[73] Finally, we will use the *Veterans SF-12 to assess health-related quality of life*. This measure has demonstrated good internal consistency and is strongly correlated with socioeconomic status and morbidities.[75]

Resource Use (Aim 2). *Intervention Costs.* Therapists will use a log to record time spent in intervention-related activities for a random 20% of all treatment days. Time records will be combined with wage data from the VA Financial Management System to estimate intervention-specific personnel costs. Technology costs of the AI-CBT program include fixed costs (e.g., software development and computer maintenance) plus variable costs (e.g., minute costs for IVR calls). One-time fixed start-up costs will be reported separately. *VA Inpatient and Outpatient Service Use* data will be obtained from the Musculoskeletal Diagnoses Cohort (MSD), a project currently underway as part of the VA Connecticut Healthcare System's CREATE. The MSD is developing validated algorithms for using VA electronic health record data to identify utilization events, comorbid conditions, receipt of opioid medications, and pain screening results, for patients with pain-related diagnoses. Information on non-VA admissions will be collected by the patient survey. To mitigate recall bias, we will use a two-time frame method that asks about utilization over the past 6 months and over the past 2 months with more weight given to the shorter timeframe.[76]

Treatment Satisfaction and Engagement (Aim 3). For patients in the AI-CBT group, we will calculate *IVR adherence* as we have in the past,[49] i.e., as the proportion of days during which an assessment was attempted in which one was successfully completed and the number of weeks during which the patient completed at least four out of seven requested IVR reports. Participants' judgments of *Treatment Credibility* will be assessed using a reliable questionnaire adapted from Borkovec and Nau.[77] Treatment credibility has been shown to be significantly associated with treatment satisfaction, engagement in treatment, and number of sessions attended. The *Pain Treatment Satisfaction Scale* of the Patient Outcomes Questionnaire will be used to assess patient satisfaction with various domains of pain care.[78] This 5-item measure shows good internal consistency and significant associations with staff and patient ratings of patient improvement. *Attendance in "Live" Telephone CBT Sessions and Program Dropout* In order to understand reasons for treatment dropout, we will attempt to reach samples of patients with low levels of engagement for qualitative interviews. Participants will rate their *Continued Skill Use* at follow-up for each of the target behaviors emphasized in the CBT program on a 0 (not at all accomplished) to 10 (completely accomplished) scale. As described above, *Daily IVR calls* will be used to collect data in the AI-CBT condition regarding pedometer measured step-counts, CBT skill practice, and pain-related functioning using pre-recorded questions we have used successfully in our prior studies.

Interim Data Analysis

There are currently no plans for interim data analysis.

Handling Missing Data

Approach to Missing Data. If more than 15% of a covariate is missing, we will use multiple imputation methods based on the SAS MI Procedure. We will check if the pattern of missingness is monotonic (i.e., if patients missing data at 3 months also have missing data at 6 months) and use a Markov chain Monte Carlo method that assumes multivariate normality to impute missing values. We will impute five sets of data and combine the imputed results using the SAS MIANALYZE procedure to obtain a valid estimate of the confidence limit and treatment effect.[85] When data are missing for items within scale scores, we will use recommended imputation procedures rather than deleting patients list-wise from the analysis.

Definitions of Covariates

Demographics and Other Covariates Measured at Baseline. We will measure patients' baseline *Sociodemographic and Pain Characteristics* that have been shown to be associated with treatment outcomes, e.g., age, gender, education level, racial/ethnic background, marital status, occupational status, pain duration, and number and location of pain sites. We also will gather data on participants' level of health literacy.[79] *Pain Classification* will be derived through a systematic evaluation of each enrolled participant's EMR using an assessment tool based on clinical guidelines for diagnosing and treating low back pain.[1] Pain classes will include non-specific low back pain, low back pain with a radicular component, or low back pain associated with other specific spinal causes. A nurse practictioner, who is supported by the West Haven COIN and who has been trained to use this tool and is using it in our trial IIR 09-058, will conduct the assessment. She will be supervised by a pain medicine physician (Dr. Rosenberg) who will review ten percent of the classification ratings for accuracy and provide corrective feedback if necessary. *Psychiatric and Substance Abuse Comorbidities* will be measured using medical record diagnoses and mental health encounters. Additional self-report information will be collected using subscales of the Mini International Neuropsychiatric Interview (MINI)[61] related to mood and substance abuse disorders. *Pain Medication Use* will be assessed through patient survey and a review of computerized pharmacy records. Pain medication will be coded as non-steroidal anti-inflammatory, non-narcotic analgesics, narcotic analgesics, and benzodiazepines and other sedative/hypnotics. For each category, we will document post-treatment whether patients have experienced an increase, no change, or decrease in their medication use. *Distance from VA* will be calculated and used as a measure of geographic access. The *Pain Catastrophizing Scale* is a 13-item self-report scale that examines thoughts and feelings people may experience when they are in pain including rumination, magnification, and helplessness.[80] Finally, *Pain-related Fear* will be measured using the Tampa Scale of Kinesiophobia-revised (TSK-R), which has two subscales (Fear of Harm/Activity Avoidance and Pathophysiological Beliefs) and has been shown to be sensitive to treatment-related change.

Methods for Dealing with Data Transformations

We will examine the distribution of all study variables to assess extreme values, missing data, variances, skewness, and type of distribution. Preliminary analyses will examine the distributions of variables to assess deviations from normality and other deviations from assumptions of the planned statistical analyses. Outliers will be corrected. Transformations of the data will be conducted if appropriate to obtain normality of residuals.

Definitions of the Analytical Sets

Baseline Comparability, Reach, and Representativeness. We will examine baseline differences across

groups in measures of study endpoints as well as other potential prognostic indicators, such as patients' age, comorbid diagnoses, and history of pain treatment. Any differences across groups in baseline characteristics will be controlled statistically in analyses comparing outcomes. The RE-AIM framework is a methodology for systematically considering all strengths and weaknesses of an intervention in order to better guide program planning.[83] To evaluate reach, we will ask patients who decline study participation whether they would be willing to provide informed consent to participate in a brief survey that identifies their reasons for declining participation and the characteristics that differentiate them from enrollees.

Analysis of Endpoints (addressing Specific Aim 1). We will test for non-inferiority of AI-CBT compared to standard telephone CBT at 12 weeks by comparing the upper limit of the two-sided 95% confidence interval for the difference in the mean RMDQ scores, calculated as AI-CBT minus standard CBT, to the pre-set non-inferiority margin of 2 points in RMDQ and will conclude that AI-CBT is non-inferior if the upper limit is less than 2. Because intent-to-treat analysis can raise the risk of type I error in a non-inferiority trial,[81] we will conduct both a per protocol and intent-to-treat analysis. The "per protocol" group assignment will be defined as completing four or more CBT sessions (either IVR or "live" sessions); however, we will revisit this definition with the Expert Panel prior to the start of the trial. We will declare AI-CBT non-inferior to standard CBT only if AI-CBT is shown to be non-inferior using both the intention-to-treat and per-protocol analysis sets. We expect that RMDQ scores will be normally distributed. If not, we will use transformations to achieve normality. We also will develop a two-level linear mixed-effects model that uses RMDQ follow-up scores at both assessment times (three and six months) as the dependent variables; treatment group, time and the treatment by time interaction as categorical explanatory variables; and baseline RMDQ score as a continuous covariate. If this model shows no significant time by group effect, we can drop the interaction term and test for the time-averaged non-inferiority of AI-CBT compared to standard CBT between 3 and 6 months. This model will also allow for adjustment for design-related factors (e.g., site and age). An unstructured variance-covariance matrix will be used to model the error variance. Secondary outcomes including pain intensity, emotional functioning, global perception of change, and quality of life will be analyzed in a manner similar to that used for the primary outcome. Analyses of all other outcomes will be conducted on an intent-to-treat basis.

Intensity of Service Use (Specific Aim 2). We will compare service utilization by category (e.g., CBT therapist time, PCP visits, and pharmacy use) between groups. We will conduct a budget impact analysis[84] and will include the cost of the intervention (personnel, supplies, CBT therapist training, and IVR fixed/variable costs) as well as costs for specific medical care services likely to be affected. Data from CBT therapists time records will be combined with wage data from the VA Financial Management System to produce estimates of intervention-specific personnel costs. Costs associated with the use of specific medical care services, such as medications, will be obtained from the Decision Support System (DSS) files. Cost analysis will be conducted in accordance with the guidance provided by Mauskopf et al.[84] including the use of sensitivity analysis and scenarios that allow for varying assumptions about intervention uptake, compliance or component costs. All resource use and cost comparisons will be adjusted for any observed differences in baseline characteristics. Because costs of resource utilization are usually skewed, alternative modeling techniques (e.g., log-transformed costs, negative binomial regression) will be used.

Intervention Engagement and Satisfaction with Care (Specific Aim 3). As in our prior research,[49] we will conduct extensive analyses of the process of intervention delivery in both arms. We will monitor the proportion of telephone CBT sessions that are completed, and we will determine the patient and session characteristics associated with patients' reports of skill practice. Patients in the AI-CBT group will report their satisfaction with aspects of the intervention (e.g., whether it provided information useful for achieving behavioral targets), and we will assess the correlation between satisfaction ratings and measures of: intervention engagement, patients' baseline characteristics, and changes in pain-related functioning. Differential dropout across groups will be examined using Kaplan-Meier curves and survival models.

Preplanned Subgroup Analysis. Because AI-CBT will continue to learn patterns in patients' experience throughout the intervention period, we hypothesize that the second 50% of patients randomized will show

an even larger difference in clinician time than the first 50%, while still demonstrating non-inferiority in pain-related outcomes. Differences in pain related functioning and in clinician treatment time across treatment groups will be tested in this subgroup analysis after stratifying the sample into early versus later recruits.

Mediators and Moderators of Intervention Effects. We will use multivariate modeling to identify the mechanisms through which the intervention achieves effects on outcomes and differential effects across subgroups.[86] Initial models will include only treatment group as the predictor. Subsequent nested models will introduce potential mediators (such as the number of completed therapist sessions), and we will evaluate changes in the relationship between experimental condition and outcomes before and after covariates are introduced. Analyses of effect moderation will focus on baseline pain severity and comorbid diagnoses using standard approaches to evaluate interactions between these covariates and patients' experimental condition.[86] Significant interactions will be interpreted by plotting regression lines for predicted outcomes of patients with high and low values of the moderator.

Evaluating the Reliability of Patients' IVR Reports. We will evaluate the integrity of IVR-reported step counts and functioning by examining associations between IVR reports and baseline characteristics that the literature suggests would be associated with patients' functioning (e.g., baseline SF-12 scores, comorbid medical diagnoses, and age). We also will examine serial correlations across IVR reports under the assumption that all correlations between scores and proximal scores should be positive and roughly of equal magnitude controlling for the time difference between reports.

RE-AIM[83] Dimensions of Intervention Adoption, Implementation, and Maintenance. Adoption will be evaluated by examining variation in study participation and intervention engagement across sociodemographic and clinical subgroups of eligible patients. For example, we will determine whether older patients or those with less education have more difficulty responding to queries about their step counts or pain-related functioning via IVR. Adoption at the provider level will be monitored by recording the proportion of providers who are willing to have their patients participate and providers' reasons for not participating. Implementation and maintenance will be evaluated through semi-structured questions at follow-up designed to identify program characteristics that might be a barrier to patients' use of the intervention in other settings and the intervention characteristics that patients feel would make it more valuable to others with chronic pain. We also will meet with clinicians in each site to gauge their willingness to adopt and maintain a similar intervention, and the ways such a system can be designed to best complement existing services.

Qualitative Interviews and Mixed-Methods. We will use audio-taped interviews with patients, CBT therapists, and PACT team members to provide a context for interpreting intervention effects and suggest additional subgroup analyses. Interviews will be transcribed verbatim, and we will enter the transcripts into NVivo for file storage and selective retrieval. Using accepted techniques,[87] Drs. Piette and Heapy will independently read transcripts, approaching the data with analytic categories in mind, but identifying other categories in the data. An iterative process will be used until agreement is reached on categories and their definitions, after which we will develop a coding template and enter it into NVivo as a tree diagram.

List of Adverse and Serious Adverse Events to be Monitored

We will track all events which our local IRB defines as potentially adverse.