STATISTICAL ANALYSIS PLAN

ClinicalTrials.gov Protocol Identifier: NCT04523714

Tailored Non-Pharmacotherapy Services for Chronic Pain: Testing Scalable and Pragmatic Approaches (RESOLVE Study)

Study Principal Investigator:
Lynn L. DeBar, PhD, MPH
Kaiser Permanente Center for Health Research (KP CHR)

Study Lead Statistician:
Andrea J. Cook, PhD
Kaiser Permanente Washington Health Research Institute (KPWHRI)

Funded by: National Institute on Aging (NIA) as part of the
Helping to End Addiction Long-term (HEAL) Pain Effectiveness Research Network (ERN)

Version Number: 6.0

Version Date: 18 July 2024
Based on Study Protocol Version 5.0; 10 November 2022

Summary of Changes from Previous Version:

Date	Description of Change (Include specific sections affected)	Rationale for Change
08/06/20	First approved version dated March 30, 2020 was erroneously labeled version 0.4 (draft numbering); it should have been 1.0. Revised SAP labeled 2.0 version.	Follow proper versioning to ensure accurate tracking of SAP
08/06/20	Corrected the phase description of clinical trial to phase III (pages 2 and 15); it was incorrectly stated as phase IV.	Inaccuracy in text
08/06/20	Made minor clarifications in section 1.0 Protocol Summary related to process and timing for randomization notification.	Clarify and ensure accuracy of description
08/06/20	Made minor edits to wording of Specific Aims: how interventions are referenced (web-based vs. online and virtual coach-led vs. telephonic); distinguish secondary from exploratory aims in aim 1a.	Align with final protocol
08/06/20	Removed incorrect description of screening as “two-step” from section 1.2	Inaccuracy in text
08/06/20	Updated section 1.3 to align with protocol’s description of interventions; changes to wording and not substantive. Made changes to how interventions are referenced (web-based vs. online and virtual coach-led vs. telephonic) throughout the document, as needed.	Clarify and ensure accuracy of description
08/06/20	Corrected modified BPI-SF description in section 3.0 to be 11-item (vs. 10) and 4 pain intensity items (vs. 3)	Inaccuracy in text
08/06/20	Added description of subgroup analyses based on age and race/ethnicity, as required for NIH phase III clinical trial (sections 4.0, 13.0 and 16.0)	Required by NIH and requested by NIA
08/06/20	Added final info on perceived support measure to mediator table in section 4.0	Finalized based on pilot study completion
08/06/20	Updated time points for assessment of mediators to include baseline or randomization notification and 3 months only	Align with final protocol
08/06/20	Updated inclusion and exclusion criteria to align with final criteria in protocol in section 6.0	Finalized based on pilot study completion
08/06/20	Changed randomization stratification based on sex to be based on EHR data and not self-reported sex in baseline survey in section 8.0	Changed due to ability to ensure EHR variable is not missing
08/06/20	Changed description of coding of relatedness of SAEs in section 19.3 to: 1) related, 2) possibly related, 3) not related (it was: definitely, probably, possibly, definitely not	The 3-level coding specified is how it is described in DSMP
09/18/20	Added rural/medically underserved residency as stratum for randomization in section 8.0 and adjusted variable in section 14.0	In response to DSMB recommendation
09/18/20	Added assessment of MCID for two subscales of pain severity composite measure (MCID in pain intensity and MCID in pain-related interference) in sections 3.0 and 15.0	In response to DSMB recommendation
09/18/20	Modified analysis plan to handle clustering within health coach in sections 14.0 and 15.0	In response to DSMB recommendation
09/25/20	Specified one additional EHR-based inclusion criterion to be applied at KPWA site only: no opioid use disorder (OUD) ICD-10 codes in section 6.1	To ensure coordination with another HEAL study at KPWA site specifically treating those with pain and OUD

10/19/20	Corrected location of Essentia Health (not Iowa but Wisconsin) in section 1.0	Error in description
10/19/20	Removed physician chart review of hospitalizations at each site as part of systematic/routine SAE identification in EHR data. DSMB approved change via email. Aligned description of ad hoc SAE identification and review with DSMP in section 19.2	Hospitalization review was modification from original, peer-reviewed DSMP; had not been planned, budgeted or deemed necessary for this minimal risk study.
07/01/21	Modified EHR-based inclusion criteria related to the minimum number of pain-related encounters (criterion #4 in section 6.1) for the 3 KP sites; changed to ≥ 2 encounters that are at least 60 days apart. Essentia remains at least 1 encounter. 6.1 Participant Inclusion and Exclusion Criteria	High rate of refusal prior to screening due to no chronic pain and ineligibility among those screened at KP sites. By increasing threshold for required pain visits, more likely to efficiently identify people likely to have chronic pain.
07/01/21	Added an EHR-based exclusion criterion related to surgery in the past 60 days for all sites 6.1 Participant Inclusion and Exclusion Criteria	High rate of refusal and reporting recent surgery and therefore no current chronic pain. By adding criterion, more likely to efficiently identify people likely to have chronic pain.
07/01/21	Added exclusion criterion related to planned/scheduled surgery in next year. 6.1 Participant Inclusion and Exclusion Criteria	Multiple virtual health coach intervention participants have indicated upcoming surgery; common exclusion criterion for pain studies
03/29/22	Revised description of active interventions throughout document to clarify painTRAINER as “CBT-based.” This affected wording in many places. Most significantly, it changed Aim #1 wording from this: **Aim #1:** Determine the effectiveness of a web-based CBT-CP program and virtual coach-led (telephonic/video) CBT-CP on achieving clinically meaningful improvements in patients’ pain severity (pain intensity + pain-related interference) relative to those receiving usual care at 3 months. To revised: **Aim #1:** Determine the effectiveness of an online, CBT-based pain management program and virtual coach-led (telephonic/video) CBT-CP on achieving clinically meaningful improvements in patients’ pain severity (pain intensity + pain-related interference) relative to those receiving usual care at 3 months.	NIH Project Scientist, Dr. Wandner, requested edit; using the consistent description of painTRAINER as CBT-based and not CBT was agreed upon for all HEAL studies that include painTRAINER.
03/29/22	Removed references to painTRAINER as an “application” and replaced with “online program” as it is not an application (“App”) that can be downloaded to a mobile device but can only be used online via a web browser. Also removed “web-based” as descriptor and replaced with “online” as it is clearer to describe as “online CBT-based.” (vs. web-based, CBT-based)	“Application” was inaccurate description.
03/29/22	Added information on sensitivity analysis related to population randomized after implementation of “enhanced” painTRAINER onboarding guidelines. [Section 14, page 20]	Requested by DSMB.
07/17/24	Corrected Dr. DeBar’s institution on the title page which changed from KP CHR to KPWHRI. This was addressed in the modification of the Protocol to version	Dr. DeBar transferred KP institutions (as of Jan. 1, 2023).

	5.0 (Nov. 2022). The protocol version info on the title page has also been updated to reflect this.	
07/17/24	Two corrections were made to the descriptions of outcomes in sections 3.2 and 3.3 to reflect the final protocol. This is not a change to the actual outcomes assessed, it is only a change to this SAP document which had 2 errors related to outcomes' descriptions. Specifically, in section 3.2, the description of the Guy/Farrar Patient Global Impression of Change stated "(1 item)", however as the Protocol accurately states, it should say "(1 item each for pain status and overall status)," i.e., 2 items. This has been corrected. In section 3.3, the exploratory outcomes related to "Examining the impact of the active interventions on high impact chronic pain and graded chronic pain" had been erroneously omitted from this document, although they are clearly stated in the study aims. This has also been corrected.	Error in descriptions.
07/17/24	Added inverse probability of missing weights (IPW) to address missing data for those missing all time points. Current missing data approach imputes data using pattern mixture approach only among those with at least one follow-up time point. Adding IPW allows for estimates to be interpreted among all randomized and not only among those with at least one follow-up time point.	Due to observed higher than expected rates of missing all outcomes across all time points (12.1%) this improvement in handling missing data was proposed by the masked Biostatistician, Dr. Andrea Cook, who does not have access to the outcome data nor has seen any summary of data by treatment arm at the time of this modification.
07/17/24	Clarified in Section 9.0 on Blinding that masking occurs at 2 levels: treatment assignment and outcomes data. Added info regarding blinding to outcomes data.	Clarification requested by NIA Project Officer and NIA's Behavioral and Social Clinical Trials Office in 07/07/24 email to Dr. DeBar

SIGNATURE PAGE AND APPROVALS

I acknowledge that I have reviewed and approved this document.

Lynn DeBar, PhD, MPH *Study Principal Investigator*	Signed by: Lynn DeBar F7B1F9BB5FAA4B4...
Andrea Cook, PhD *Study Lead Statistician* *KPWHRI*	Signed by: Andrea Cook 614B9C07F33F4AE...
Richard Thompson, PhD *Heal Resource Center Statistical Core* *Johns Hopkins University*	DocuSigned by: 213038F45C744F7...
T. Charles Casper, PhD *Heal Resource Center Data Coordinating Center* *University of Utah*	Signed by: T. Charles Casper 227A234507EE43C...

Table of Contents

List of Abbreviations:

AE	Adverse Event
CBT-CP	Cognitive Behavioral Therapy for chronic pain
CRF	Case Report Form
DCC	Data Coordinating Center
DSMB	Data Safety Monitoring Board
EDA	Exploratory Data Analyses
EHR	Electronic Health Records
eCRF	Electronic Case Report Forms
GCP	Good Clinical Practice
HIPAA	Health Insurance Portability and Accountability Act
IRB	Institutional Review Board
ITT	Intention-To-Treat
KP	Kaiser Permanente
LSMEANS	Least-squares Means
MCID	Minimal Clinically Important Difference
MedDRA	Medical Dictionary for Regulatory Activities
MOP	Manual of Procedures
NCT	National Clinical Trial
NIH	National Institutes of Health
NIH IC	NIH Institute or Center
OHRP	Office for Human Research Protections
PI	Principal Investigator
QA	Quality Assurance
QC	Quality Control
SAE	Serious Adverse Event
SAS	Safety Analysis Sample
SAP	Statistical Analysis Plan
SDOH	Social Determinants of Health
SOP	Standard Operating Procedure
UC	Usual Care
UP	Unanticipated Problem
US	United States

1.0. Protocol Summary

The RESOLVE study is a phase III, randomized, comparative effectiveness trial evaluating 2 Cognitive Behavioral Therapy (CBT)-based chronic pain treatments delivered via different telehealth modalities: 1) online program and 2) live, coach-led virtual sessions (telephone and/or video-administered). The study uses a 3-arm, parallel intervention design; both intervention arms will be compared to usual care services.

Participants will be recruited from the populations of 4 integrated health care systems which serve as the clinical sites: 1) Kaiser Permanente Georgia (KPGA), 2) Kaiser Permanente Northwest (KPNW), 3) Kaiser Permanente Washington (KPWA) and 4) Essentia Health (located in Minnesota, North Dakota, and Wisconsin).

Potentially eligible participants will be identified based on electronic health records (EHR) criteria and then mailed information on the study and an invitation to complete eligibility screening by phone or web. Individuals who meet screening eligibility criteria will complete the informed consent process by phone or web and then the baseline survey, ideally during the same phone or web encounter. Upon completion of the baseline survey, individuals will be randomized within the study's electronic data capture system and notified of their study arm assignment by mail within approximately one week.

Participants randomized to either of the two intervention arms will complete 8 sessions online or by phone/video; those randomized to usual care will receive an educational resource manual for pain management. All study participants will be enrolled for 12 months from the time of randomization and complete self-reported assessments at baseline (prior to randomization), notification of randomization (approximately one-two weeks from randomization), and 3-, 6- and 12-months from randomization.

1.1. Specific Aims

Aim #1: Determine the effectiveness of an online, CBT-based pain management program and virtual coach-led (telephonic/video) CBT-CP on achieving clinically meaningful improvements (≥ 30%) in patients' pain severity (pain intensity + pain-related interference) relative to those receiving usual care at 3 months from randomization.

1a. Examine the impact of the active interventions on secondary pain outcomes and related quality of life outcomes (social role functioning, physical functioning, and patient global impression of change); as well as exploratory outcomes, which include long-term opioid use, comorbid symptomology (depression, anxiety, and sleep disturbance), and high impact and graded chronic pain.

1b. Conduct subgroup analyses to determine the impact of the active interventions on specific populations and explore for potential heterogeneity of treatment effects by sex; rural/medically underserved residency; multiple pain conditions; mental health mood disorders; and negative social determinants of health.

1c. Examine the role of theory-based mediators, pain catastrophizing, pain-related self-efficacy, and perceived support, on pain-severity.

Aim #2: Assess the cost and incremental cost-effectiveness of the online and virtual coach-led CBT-CP interventions compared to each other and usual care.

Aim #3: Conduct a qualitative evaluation to understand: 1) patient experiences of the interventions, including how they relate to treatment response, variability by site, and rural/medically underserved residency status; and 2) health system issues, including

adaptations and contextual factors at the site and external levels, barriers and facilitators to intervention success and potential for adoption, sustainability and dissemination.

NOTE: The analytic approaches for Aims #1-2 are the focus of this Statistical Analysis Plan (SAP). Aim #3 is a qualitative evaluation and does not utilize statistical analyses; therefore, it is not addressed in this SAP.

1.2. Study Population

The 2,331 participants we will randomize (777 per study arm) will include those who receive primary care services from 4 integrated health care systems/clinical sites: KPGA, KPNW, KPWA and Essentia Health and who have high-impact chronic pain as demonstrated by electronic health records (EHR)-based eligibility criteria and a screening assessment.

1.3. Description of Interventions

Two interventions will be tested in this comparative effectiveness trial.

1) An online, CBT-based pain coping skills training program (painTRAINER)
2) A virtual, live, coach-led CBT-CP intervention (telephone and/or video-administered)

Both interventions are comprised of 8 sessions which participants are expected to complete weekly. The content of the 2 interventions is analogous and focuses on interactive training in evidence-based pain-coping skills.

Our analysis will include an economic evaluation comparing the costs, and incremental cost-effectiveness of the active interventions. We expect that the live coach and other elements of the virtual, coach-led intervention will make it more expensive (and possibly more effective) than the online intervention (painTRAINER).

Those randomized to the usual care arm of the study will be sent a bound copy of the 2020 edition of the *American Chronic Pain Association Resource Guide to Chronic Pain Management*.

All enrollees can receive any pharmacologic (prescription or over the counter) and nonpharmacologic treatments available to them without restriction.

2.0. Introduction

This Statistical Analysis Plan (SAP) provides the details of statistical considerations, analyses, and reports planned for the RESOLVE study, including proposed regression models, sample size estimates, power considerations, and safety analyses. The proposed analyses will be conducted on the entire participant sample and on pre-specified subgroups, as described in this SAP. In addition, this plan discusses statistical issues relevant to these analyses (e.g., sample data to be used, missing data, adjustments for multiplicity, etc.).

For the HEAL Pain ERN, the Data Coordinating Center (DCC) at the University of Utah has the primary responsibility for data quality. For safety analyses and DSMB report production, the JHU-Tufts Statistical Core, physically located at the Johns Hopkins University will examine or oversee data quality, create secondary variables, and perform exploratory data analyses (EDA).

All variables in the analytic database will be evaluated to detect gaps, patterns, and inconsistencies in the data. These analyses will emphasize examination of the nature and extent of variability for all variables. Visual techniques to explore continuous variables may include histograms and boxplots. Outliers will additionally be examined for data entry errors. Summary statistics for continuous variables will include the number of patients, median, mean, standard deviation, and range, while statistics for categorical variables will include the frequency and percentage of patients in each category.

Statistical analyses will be performed on data that have been quality-assured through DCC and Statistical Core protocols and monitoring reviews, and that have been exported directly from the RESOLVE study database. The following analysis procedures may be applied to blinded (no treatment arm designation, for study investigators and staff), partially unblinded (treatment designation only as “A”, “B”, or “C” for DSMB and at interim analyses), and unblinded (full treatment designation, for final analyses) data, by Statistical Core and DCC staff having the appropriate role permissions.

All data collection procedures and statistical analyses in this SAP will be finalized and approved by the RESOLVE Study Principal Investigator (PI) and Study Lead Statistician prior to the unblinding of the data. In addition, the data collection procedures and statistical analyses detailed in this SAP may be modified and are given precedence over the analytical plans outlined in the Study Protocol; however, any modifications or changes in the primary endpoint and/or its analysis will also be reflected in a protocol amendment. Any updates to the SAP will be thoroughly detailed and documented at the beginning of this document.

3.0. Study Outcomes

3.1. Primary Effectiveness Trial Outcome

Primary Outcome	Brief Description of Measure	Outcome Measured By	Time Frame
Achieving minimal clinically important difference (MCID) in pain severity at 3 months (Yes / No)	Minimal clinically important difference (MCID) in pain severity is defined as a 30% decrease in score on modified 11-item version of the Brief Pain Inventory – Short Form (BPI-SF)[1-3] from baseline (consistent with IMMPACT consensus guidelines)[4] (binary)	Patient self-reported pain at baseline and 3 months	Baseline to 3 months

3.2. Secondary Effectiveness Trial Outcomes

Secondary Outcomes	Brief Description of Measure	Outcome Measured By	Time Frame
Achieving MCID in pain severity at 6 and 12 months (Yes / No)	See above (binary)	Patient self-reported pain at baseline, 6 and 12 months	Baseline to 6 and 12 months
Achieving MCID in pain intensity (Yes / No)	Minimal clinically important difference (MCID) in pain intensity is defined as a 30% decrease in score on 4-item pain intensity subscale of the Brief Pain Inventory – Short Form (BPI-SF)[1-3] from baseline	Patient self-reported pain at baseline, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months
Achieving MCID in pain-related interference (Yes / No)	Minimal clinically important difference (MCID) in pain intensity is defined as a 30% decrease in score on 7-item pain-related interference subscale of the Brief Pain Inventory – Short Form (BPI-SF)[1-3] from baseline	Patient self-reported pain at baseline, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months
Pain severity	BPI-SF; composite of pain intensity (4 items) and pain-related interference (7 items) subscales (11 items; continuous)	Patient self-reported pain at baseline, 3, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months
Pain intensity	Pain intensity subscale of the BPI-SF (4 items; continuous)	Patient self-reported pain at baseline, 3, 6 and 12 months	
Pain-related interference	Pain-related interference subscale of the BPI-SF (7 items; continuous)	Patient self-reported pain at baseline, 3, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months

Secondary Outcomes	Brief Description of Measure	Outcome Measured By	Time Frame
Social role functioning	PROMIS Ability to Participate in Social Roles 4A[5] (4 items; continuous)	Patient self-reported social role functioning at baseline, 3, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months
Physical functioning	PROMIS Physical Functioning Short Form 6b (6 items; continuous)	Patient self-report physical functioning at baseline, 3, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months
Patient global impression of change (PGIC)	Guy/Farrar Patient Global Impression of Change (1 item each for pain status and overall status)	Patient self-reported impression of change at 3, 6 and 12 months	*Primary:* Baseline to 3 months *Secondary:* Baseline to 6 and 12 months

3.3. Exploratory (Tertiary) Effectiveness Outcomes

Exploratory (Tertiary) Outcomes	Brief Description of Measures	Measured By	Time Frame
Long-term opioid use (Yes / No)	Opioid prescription orders or fills indicating a continuous ≥ 60-day supply during the prior 90-day period; account for prescriptions prior to 90-day look-back which carry into the assessment period (binary)	Electronic health record (EHR) opioid prescription data assessed quarterly 90-Day Periods for Developing Quarterly Variables (Baseline survey date = Day 1) Baseline = Day -90 to day -1 3 months = Day 1 to day 90 6 months = Day 91 to day 180 9 months = Day 181 to day 270 12 months = Day 271 to day 360	Baseline to 3, 6 and 12 months
Depression symptomology	Patient Health Questionnaire-8 (PHQ-8)[6] (8 items; continuous)	Patient self-reported depression symptomology at baseline, 3, 6 and 12 months	Baseline to 3, 6, and 12 months
Anxiety symptomology	Generalized Anxiety Disorder-7 (GAD-7)[7] (7 items; continuous)	Patient self-reported anxiety symptomology at baseline, 3, 6 and 12 months	Baseline to 3, 6, and 12 months
Sleep disturbance	PROMIS Sleep Disturbance – Short Form 6a[8] (6 items; continuous)	Patient self-reported sleep disturbance at baseline, 3, 6 and 12 months	Baseline to 3, 6, and 12 months

Exploratory (Tertiary) Outcomes	Brief Description of Measures	Measured By	Time Frame
Examine the impact of the active interventions on high impact chronic pain and graded chronic pain	High Impact Chronic Pain[9,10] Graded Chronic Pain Scale-Revised[10]	Patient self-reported pain intensity, duration, and pain-related disability	Baseline to 3, 6, and 12 months

3.4. Economic Evaluation Outcomes

Cost Outcomes	Brief Description of Measures	Measured By	Time Frame
Cost and incremental cost-effectiveness	Health care utilization and intervention costs will be assessed. Using the framework of cost-effectiveness, we will estimate the incremental cost per additional patient with a MCID in pain severity (30% reduction from baseline), at 12 months, and the quality-adjusted life year (QALY) gained—utilities will be estimated using the EQ-5D-5L.[11]	Healthcare utilization costs: EHR data costed using standard costing algorithms[12,13] and Medicare fee schedules Intervention costs: Process data related to all relevant resources used in the intervention delivery EQ-5D-5L: Patient self-report at baseline, 3-, 6- and 12-months	Patient health care utilization costs from baseline to 12 months

4.0. Additional Variables for Effectiveness Analyses (Moderators and Mediators)

The following **moderators** will be assessed related to the **primary outcome.**

Moderator	Definition	Data Source
Sex	Male vs. Female/Other Sex at birth as reported by subject; assessed using HEAL CDE Demographic question	Patient self-report at baseline
Age	<65 vs ≥ 65 years old	EHR data
Race/Ethnicity	White/Non-Hispanic Black or African American/Non-Hispanic Hispanic Other	Patient self-report at baseline
Rural/medically underserved residency	Urban vs. Rural/medically underserved Rural is defined as subject's resident Census Tract corresponds to US Census 2010 Rural-Urban Commuting Area (RUCA) Codes 4, 5, 6, 7, 8, 9 or 10 (https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/) Medically underserved is defined as subject's resident Census Tract corresponds to HRSA-designated primary care or mental health geographic or geographic high needs health professional shortage area (https://data.hrsa.gov/tools/shortage-area/hpsa-find)	EHR patients' geocoded data extracted at baseline
Multiple non-malignant musculoskeletal pain conditions	1 pain cluster vs. >1 > 1 pain-related ICD-10 diagnosis corresponding to more than one (>1) of the non-malignant musculoskeletal chronic pain condition clusters developed for the National Pain Strategy chronic pain condition clusters[14] 1. Back pain 3. Neck pain 4. Limb/extremity pain, joint pain and arthritic disorders 5. Fibromyalgia 6. Headache 7. Orofacial, ear, and temporomandibular disorder pain 8. Musculoskeletal chest pain 9. General pain subcategory of the Other painful conditions cluster	EHR data; diagnoses in subject's EHR extracted at baseline for prior 360 days
Mental health mood disorders	ICD-10 diagnosis for depression and/or anxiety diagnosis	EHR data; diagnoses in subject's EHR extracted at baseline for prior 360 days

Moderator	Definition	Data Source
Negative social determinants of health (SDOH)	Negative SDOH/existing need vs. No SODH need Patient endorses need in one or more of the following domains. Financial Resource Strain (1 item) Food Insecurity (2 items) Transportation/Access Needs (2 items) Housing Instability (3 items)	Patient self-report at baseline

The following **mediators** will be assessed related to **the primary outcome.**

Mediator	Brief Description of Measure	Measured By
Pain catastrophizing	Pain Catastrophizing Scale (PCS)– Short Form 6[15] (6-item; continuous)	Patient self-report at baseline and 3 months
Pain self-efficacy	Self-Efficacy for Pain Management subscale of the Chronic Pain Self-Efficacy Scale[16] (5-item; continuous)	Patient self-report at baseline and 3 months
Perceived support	2 items; developed in UG3 phase	Patient self-report at notification of randomization and 3 months

5.0. Sample Size Determination for the Primary and Secondary Effectiveness Analyses

For our primary study outcome, clinically meaningful improvement (MCID; 30% reduction in overall BPI-SF score) in pain severity at 3 months, we calculated sample size requirements for a two-sided comparison of independent proportions with 90% power using Fisher's Least Significant difference method to account for three-way comparisons (specifically conduct Omnibus Wald Test for any difference between three groups and then conduct pairwise comparisons given an overall difference is statistically significant at the 0.05 alpha level). Sample size calculations assumed a usual care outcome rate of 15% who achieve the MCID of 30% reduction in pain severity from baseline to 3 months. The 15% usual care rate was chosen because we observed this rate among participants in the usual care arm of our recently completed PPACT trial[17] and it is similar to rates observed by others in the usual care arm of similar trials.[18] We calculated the necessary sample size of 1,863 (621 per arm) to detect a difference of 7.5% between a given intervention group relative to usual care in the proportion of individuals who attain a clinically meaningful change (30% reduction from baseline) in pain severity. We estimate a retention rate of 80%. Thus, to achieve this final sample size, we will randomize 2,331 individuals, 777 per intervention arm. A 7.5% detectable difference corresponds to 22.5% (relative change of 150%) of individuals in an intervention arm attaining a clinically meaningful change in pain severity.

We also estimated power for secondary analyses related to aim 1b (subgroup analyses). We calculated sample size requirements for these secondary analyses using the same assumptions described above (i.e., a two-sided comparison of independent proportions using Fisher's Least Significant difference approach to account for three-way comparisons, with sample size calculations assuming an improvement rate of 15% in the usual care arm). We ranged our subgroup sample size between 20% and 40% of the original 1,863 sample size since this should cover the range of the subgroup sample sizes of interest.

Power for Subgroup Analyses Ranging the Size of the Subgroup

Assumptions			80% Power to Detect:		
Subgroup Sample Size	**Usual Care (UC)**		**Intervention (Int)**	**Detectable Difference**	**Relative Change**
N (% of 1,863)	% UC		% Int	%Int - %UC	% Int / %UC
372 (20%)	15.0%		31.0%	16.0%	206.7%
558 (30%)	15.0%		27.7%	12.7%	184.7%
744 (40%)	15.0%		25.8%	10.8%	172.0%

We have 80% power to detect a 10.8% difference (relative change of 172%) between each intervention group and UC if subgroup sample size is 40% of the overall study population and 16.0% difference (relative change of 206.7%) if the subgroup sample size is 20% of the overall study population.

Note that in our previous studies we observed for sex, comorbid mental health conditions (depression, anxiety) not to be less than 40% of our population suggesting that we have high power for our primary subgroups of interest. Further, we are aiming to have at least 20% of rural/medically underserved residency and therefore we have good power for this subgroup. Negative social determinants of health is an exploratory subgroup since we are not sure how many people in this population a priori will have this indication.

6.0. Definition of Study Samples

6.1. Participant Inclusion and Exclusion Criteria

This study will employ a population-based recruitment approach in which the individuals who are invited/recruited to potentially participate in the study will be identified based on the following electronic health records (EHR)-based criteria.

Participant Inclusion Criteria

Individuals must first meet these EHR-based criteria in order to be invited to be screened for participation.

1. Active/enrolled in one of the 4 participating integrated health care systems at the time of query and for the prior 360 days
2. Age 18 years or older (based on date of birth)
3. English speaking or do not need interpreter services
4. Have at least 1 [at Essentia] or at least two which are >60 days apart [at KP sites] outpatient pain-related health care encounters with nonmalignant musculoskeletal pain diagnoses [as determined by ICD10 codes for any of the following: back- neck-, limb/extremity-, joint-pain, arthritic disorders, fibromyalgia, headache, orofacial/temporomandibular pain, or musculoskeletal pain][14] within the past 360 days
5. Do not have an encounter for surgery related to common musculoskeletal pain conditions (e.g., joint replacement, spinal fusion, carpal tunnel release surgery) [as determined by CPT and/or ICD-10 codes] within the past 60 days
6. Do not have two or more separate encounters with a malignant cancer diagnosis other than non-melanoma skin cancer [as determined by ICD-10 codes] within past 60 days
7. Do not have ICD-10 code(s), CPT code(s) or department/provider encounters indicating receipt of hospice or other palliative care within the past 360 days
8. Do not have ICD-10 codes indicating severe cognitive impairment precluding participation in a behavioral/ lifestyle change program

Note: At the KPWA site only, one additional EHR-based exclusion criterion will be applied, which is: Do not have ICD-10 codes indicating opioid use disorder (OUD). This criterion is being applied because there is another HEAL study being conducted at KPWA that focuses on treating individuals with pain and OUD specifically.

Individuals who meet the above EHR criteria will be invited to respond to screening questions and must meet the following inclusion criteria:

1. Have high-impact chronic pain (as indicated by self-report of having pain on most or every day in past 3 months and pain limiting life or work activities on most or every day in past 3 months)[9]
2. Have persistent pain (as indicated by self-report PEG score of ≥ 12)
3. Be able to participate in either of the active interventions (i.e., have internet and phone access required for accessing treatments)

Participant Exclusion Criteria

Individuals screened who meet the inclusion screening criteria (and the initial EHR criteria described above) cannot endorse any of the following exclusion criteria:

1. Have received CBT for pain or pain-related psychoeducation or behavioral skills training within in the past 6 months (in-person, by phone or videoconference, or online)
2. Currently receiving or will be starting CBT for pain or pain-related psychoeducation or behavioral skills training in the next month (in-person, by phone or videoconference, or online)

3. Currently receiving or will be starting inpatient or intensive outpatient services for substance use disorder in the next month
4. Have a planned/scheduled surgery in the next 12 months related to pain condition

6.2. Intent-to-Treat (ITT) Sample

As the primary analysis, all effectiveness outcome measures are analyzed under the ITT. Under this principle, each eligible subject is analyzed according to the treatment group to which they were assigned at the time of randomization.

Modified intent-to-treat will not be employed in this protocol.

6.3. Safety Analysis Sample

All randomized subjects will be included in the primary safety analysis sample. Secondary analyses will exclude those in the intervention groups who were deemed as ineligible post-randomization and therefore not able to continue in the intervention group (e.g., dementia noted after randomization).

6.4. As-Treated Sample

The potential for cross-overs in this study is minimal; however, in the case of participants randomized to usual care who receive CBT outside of the study intervention, a per-protocol sample may be constructed (if greater than 10% cross-over from UC to receive CBT) and examined in which treatment-as-received is analyzed.

7.0. Definition of Treatment Adherence

Treatment adherence is defined as completing at least 6 of the 8 total sessions for each of the active intervention arms.

8.0. Randomization

After completion of the baseline data collection, study participants will be individually randomized in equal ratio to one of the three study arms.

Randomization will be stratified on:

Variable	Source
Sex	Sex from electronic health record (Male vs. Female or Other)
Pain severity score	Score on BPI-SF from baseline survey (< 7 vs. ≥7)
Clinical site	KPWA, KPNW, KPGA, Essentia
Rural/medically underserved residency	Residency geocode for Census Tract from electronic health record

Within each stratum, to contain concealment and balance randomization over time, a random permutated block design will be used with random variable block sizes of 3, 6, or 9 to ensure approximately equal accrual into the three study arms.

The TIN JHU/Tufts Lead biostatistician will develop the randomization scheme and the TIN Utah DCC will implement the scheme using the appropriate statistical software. The scheme to be implemented will be integrated into the study centralized tracking system.

Randomization will be conducted by KPWHRI research staff within the study centralized tracking system after consent and baseline assessment completion.

9.0. Blinding

Masking occurs at 2 levels: 1) masking to treatment assignment, i.e., not knowing which of the 3 treatment arms a participant has been assigned to in the study; and 2) masking to outcomes data, meaning not having access to follow-up study data individually, in aggregate or by treatment arm.

It is not feasible for participants to be masked to treatment arm assignment due to the type of intervention. The PI and select Co-Investigators and study staff who are involved in the administrative oversight of the study and/or delivery of the interventions will also be aware of treatment assignment but masked to outcomes data. Primary and secondary outcome assessors will be masked to treatment assignment but cannot be masked to individual outcomes data.

10.0. Multiplicity

We have one primary outcome: minimal clinically important difference (MCID) in pain severity at 3 months. To control for multiple comparisons due to 3 intervention groups we will apply Fisher's Least Significant Difference approach in which first an omnibus Wald test for any statistically significant difference between the three groups is evaluated at the 0.05 alpha level and the pair-wise differences are then evaluated, each using two-sided alpha=0.05, only if the omnibus test is statistically significant. In addition, secondary analyses will use a similar approach as that described for the primary outcome to control for the three group comparisons.

For safety outcomes, multiplicity will not be considered since this study was not powered to observe safety.

11.0. Missing Data

The clinical investigation team will make substantial efforts to ensure complete collection of data for all patients, and to ensure minimal loss to follow-up to optimize evaluation of the primary effectiveness outcome. For the primary analyses, we will analyze those patients who have follow-up data. Any missing data will be assumed to be missing at random (MAR) given the adjustment of baseline covariates. However, in the event that a patient is lost to follow-up, the underlying reason(s) for dropping out will be tracked in order to help assess the mechanisms of missingness. We will analyze missing data and conduct sensitivity analyses as required (i.e., >15% missing outcome follow-up at 3 months or differential missingness by group). Specifically, we will address missing data in two ways. Our first approach will apply a pattern mixture imputation missing data approach that relaxes the MAR assumption conditioning on the patterns of missing data over time among those with at least one follow-up time measure and inverse probability weights (IPW) for those missing all follow-up time points.[19] The second approach will be used as a sensitivity analysis assuming a worst-case, best-case approach. Specifically, for those with missing outcome data in the intervention arms we will assume they did not achieve the MCID in pain severity at each missing time point (worst-case). For those randomized to the usual care we will assume all achieved the MCID in pain severity at each missing time point (best-case). This sensitivity analysis will provide the extreme case in how small the intervention effect could be relative to usual care due to missing outcome data.

12.0. Outlier Measures

The potential for outliers is minimal in our data since most measures are from questionnaires and continuous values are scales (e.g., 11-point Likert score). Therefore, we will not correct for outliers in the analyses involving survey data. In the economic analyses we will assess for the influence of outliers with a sensitivity analysis where we Windsorize outliers to the value of the 99th percentile.

13.0. Demographic and Baseline Characteristics

Demographics will include age, sex, race, and ethnicity and will be summarized by treatment group using descriptive statistics for each of the defined samples and for the effectiveness analyses. Baseline variables are defined as pre-randomization variables such as baseline function or symptom score, comorbidities or any other variables thought to be associated with primary or secondary outcomes, moderate intervention (subgroups) or predict drop-out. Candidate variables will include but are not limited to age, sex, race, ethnicity, rural/medically underserved residency, multiple pain conditions, mental health mood disorders, and negative social determinants of health (SDOH).

Descriptive statistics for all follow-up data will be provided by intervention arm. These descriptive statistics will include the mean, median, SD, maximum and minimum for continuous variables, and frequencies, percentage, and tabulations for categorical variables. Summary statistics will be performed on patient demographics, including age, sex, race, ethnicity, and other clinical characteristics, including medical history, and prior and concomitant interventions.

Subgroups defined by cut points are: 1) sex (male), 2) age (≥65 years), 3) race/ethnicity (White/Non-Hispanic, Black or African American/Non-Hispanic, Hispanic, Other), 4) rural/medically underserved residency, 5) multiple pain conditions (>1 pain condition based on ICD dx), 6) mood disorders (anxiety or depression based on ICD dx), and 7) negative/poor SDOH. Additional analysis of groupings of these factors will be dictated by group size and balance between groups.

14.0. Primary Effectiveness Analyses

Primary Outcome Statistical Hypothesis: Both active interventions (painTRAINER and virtual coach-led CBT-CP) are expected to result in a higher proportion of patients achieving MCID (30% reduction in overall BPI-SF score) in pain severity relative to those receiving usual care at 3-month follow-up (primary time point).

Secondary Time Point Statistical Hypotheses: Both active interventions (painTRAINER and virtual coach-led CBT-CP) are expected to result in a higher proportion of patients achieving MCID (30% reduction in overall BPI-SF score) in pain severity relative to those receiving usual care at the 6 and 12-month follow-ups (secondary time points).

Primary Outcome and Secondary Time Point Analyses: We will use modified Poisson regression[20,21] fit using generalized estimating equations (GEE) to model the binary primary outcome, MCID in pain severity (30% reduction from baseline), 3 months (primary time point) and 6 and 12 months (secondary time points). We are employing modified Poisson (i.e., Poisson family with log link, but use robust standard errors to correct for mis-specified outcome variance) instead of logistic regression since the binary outcome is not rare and the estimate of interest is the relative risk. We will use a working independence correlation matrix and will calculate standard errors using the robust sandwich estimator to account for within-person and within-health coach correlation[21,22] and account for the mis-specified mean-variance structure when using Poisson regression for a binary outcome.[20,23] We will include interactions between each intervention and indicators of time (3, 6, 12 months) to estimate time-specific intervention effects; the primary comparisons will be between the interventions and usual care at 3 months (i.e., primary effectiveness will test the size of the intervention coefficient at the 3-months timepoint). We will adjust for baseline levels of pain severity, other stratification variables (sex, clinical site, and rural/medically underserved residency), and a priori variables predictive of outcome (multisite pain and co-occurring mental health condition). Specifically, we will fit the following mean model where Usual Care and 3 months are the reference groups:

$$\log\big(E(Y_{ij})\big) = \beta_o + \beta_1 Int1_i + \beta_2 Int2_i + \beta_3 T6_{ij} + \beta_4 T12_{ij} + \beta_5 Int1_i T6_{ij} + \beta_6 Int2_i T6_{ij} + \beta_7 Int1_i T12_{ij} + \beta_8 Int2_i T12_{ij} + \boldsymbol{\beta}_z \boldsymbol{Z}_i$$

where Y_{ij} is the binary outcome for participant *i* (*i*=1,…,n) and time point *j* (*j*=1,2,3), $Int1_i$ is 1 if the participant *i* is randomized to the first intervention group and 0 otherwise, $Int2_i$ is 1 if participant *i* is randomized to the second intervention group and 0 otherwise, $T6_{ij}$ is 1 if outcome is measured for participant *i* at 6-months (*j*=2) and 0 otherwise, $T12_{ij}$ is 1 if outcome is measured for participant *i* at 12-months (*j*=3) and 0 otherwise and $\boldsymbol{Z}_i$ is a vector of baseline adjustment covariates.

In addition, we will conduct a sensitivity analysis to address a change in study procedures that occurred after the study launched. Specifically, to improve engagement in the painTRAINER intervention the study team added motivational interviewing-based language to the onboarding script for those randomized to the online program. This change was implemented on 8/9/2021 after ~140 patients were enrolled in the study. The sensitivity analysis will repeat the primary analysis amongst only those who were randomized after 8/9/2021, this allows to evaluate whether after implementation of the "enhancement" to the painTRAINER engagement activities, there may be a stronger intervention effect.

All analyses will be conducted following an intent-to-treat (ITT) approach, including all individuals randomized regardless of their engagement with or exposure to the intervention.

15.0. Secondary Effectiveness Analyses (Secondary Outcomes)

Secondary Outcomes (Aim1a) Primary Statistical Hypotheses: Those in both active intervention conditions will show greater improvements in overall pain severity, pain intensity (MCID and overall), pain-related interference (MCID and overall) and related quality of life outcomes (social role functioning, physical functioning, and patient global impression of change) relative to those receiving usual care at the 3-month follow-up.

Secondary Outcomes (Aim1a) Secondary Statistical Hypotheses: Those in both active intervention conditions will show greater improvements in overall pain severity, pain intensity (MCID and overall), pain-related interference (MCID and overall) and related quality of life outcomes (social role functioning, physical functioning, and patient global impression of change) relative to those receiving usual care at the 6- and 12-months follow-ups (secondary time points).

Secondary Outcomes (Aim 1a) Analyses: Aim 1a will examine the impact of the active interventions on secondary outcomes. Analyses of these outcomes will follow the same proposed approach as Aim 1. We will use linear regression for continuous outcomes and Poisson regression for binary and count outcomes. We will use GEE [21,22] to estimate regression models for longitudinal data using an independence working correlation matrix. We will calculate all standard errors using the robust sandwich estimator[21,22] to account for within-person and within-health coach correlation or any mis-specified variance structures. We will include an interaction between intervention arms and time indicators, and the primary time point for all secondary analyses will be 3 months following randomization and will include as covariates baseline levels of pain severity and all stratification variables.

16.0. Additional Effectiveness Analyses for Primary Outcome (Moderators and Mediators)

Moderator (Aim 1b) Statistical Hypotheses: We hypothesize that for patients with more complex conditions (concomitant mood or anxiety disorders; multisite pain) or challenging social / environmental factors (rural/medically underserved or unmet social needs) the contact with a live coach will result in better pain-severity outcomes when compared to those receiving painTRAINER.

Moderator (Aim 1b) Analyses: Subgroup analyses will be conducted to determine the impact of the active interventions on specific populations and explore for potential heterogeneity of treatment effects by sex, age, race/ethnicity, rural/medically underserved residency, multiple pain conditions, mental health mood disorders, and negative social determinants of health. Analyses for Aim 1b will follow the same general approach as for Aim 1 but will be focused on assessing heterogeneity of treatment effects (sub-groups). We will assess heterogeneous treatment effects by each potential moderator separately. For each moderator, we will include in the regression models described in the analytic plan for Aim 1, a main effect for the moderator and an interaction between the moderator, intervention, and follow-up time, to estimate time-specific intervention effects within each subgroup defined by the potential moderator. The primary comparison for Aim 1b will be of the interaction terms associated with each intervention arm at the 3-months follow-up time. The longitudinal nature of data collection will allow us to qualitatively assess if the treatment effectiveness pattern is different over time in each of the intervention groups at the two levels of each of the moderators.

Mediator (Aim1c) Statistical Hypotheses: We hypothesize that changes in pain catastrophizing, pain-related self-efficacy, and perceived support from baseline to 3 months will be mediators of treatment outcomes at 6 months for those in both active interventions.

Mediator (Aim 1c) Analyses: We will use mediation analyses to assess and quantify the effect of theory-based mediators (pain-catastrophizing, pain self-efficacy, perceived support). Mediators represent a causal pathway between the intervention and outcome. Mediation occurs when the intervention influences a variable (the mediator) that in turn subsequently influences the outcome variable. Controlling for a mediator variable causes the strength of relationship between intervention and outcome to be meaningfully reduced. Consistent with recommendations, we will conduct mediation analyses only for interventions that have significant impacts on the outcomes under consideration at 6 months. Primary mediation analyses will assess the effect of the potential mediators on the primary outcome at 6 months, MCID in pain severity, while explanatory secondary analyses will investigate mediator impacts on secondary outcomes at 6 months. We will conduct mediation analyses using the framework recommended by Baron and Kenney[24] but using more recent statistical methods developed to better quantify and decompose different aspects of the mediation effect.[25] We will run a separate set of mediator analyses for each intervention compared to usual care. We will illustrate our approach for the binary outcome, achieving MCID in pain severity at 6 months.

Our statistical approach for mediation analysis will be to first demonstrate an association between the interventions and the binary outcome, achieving MCID in pain severity at 6 months (step 1; Aim 1).

Step 1 Model:

$$\log(E(Y_i)) = \beta_o + \beta_1 Int_i + \boldsymbol{\beta}_z \boldsymbol{Z}_i$$

where Y_i is the binary outcome achieve MCID in pain at 6 months for participant *i* (*i*=1,…,n), Int_i is 1 if the participant *i* is randomized to given intervention group and 0 otherwise and $\boldsymbol{Z}_i$ is a vector of baseline adjustment covariates.

We will then demonstrate an association between the interventions and change in each mediator at 3 months (step 2). To do this, we will construct a regression model for each potential mediator under consideration, with the change between the value of the mediator recorded at 3 months and baseline as the dependent variable including a main effect for the intervention. We will do this for each potential mediator and include in the next step only those mediators that are associated with either intervention group with a p-value less than 0.10 (test based on β_{M1} below).

Step 2 Model:

$$E(M_i) = \beta_{Mo} + \beta_{M1} Int_i + \boldsymbol{\beta}_{Mz} \boldsymbol{Z}_i$$

where M_i is the change in mediator of interest from baseline to 3 months for participant *i* and we will fit this model using linear regression.

The next step (step 3) will be to demonstrate the reduction of the intervention effect on the outcome after removing the effect of the mediator(s). To do this, we will construct a multi-mediator inverse probability weighted (IPW) regression model. This approach allows us to estimate the direct effect of the intervention after rebalancing the intervention and UC groups with respect to the mediators. The application of the IPW approach, as compared to the traditional approach of adjusting for multiple mediators, will allow us to more appropriately account for confounding between a mediator and the outcome both by additional mediators and by other measured variables. Specifically, we will first model the probability of receiving the active intervention given the change in mediators from baseline to 3 months (all mediators that were found associated with intervention in step 2) using logistic regression adjusting for potential baseline confounders (Step 3a).

Step 3a Model:

$$logit(E(Int_i)) = \beta_{Wo} + \boldsymbol{\beta}_{WM} \boldsymbol{M}sub_i$$

where $\boldsymbol{M}sub_i$ is vector of the subset of mediators that are shown to be associated with the intervention group in Step 2. This model will be used to construct weights defined as the inverse of the estimated probability that each person was assigned to their intervention given their observed mediator value. Specifically, the weights will be:

$$w_i = \begin{cases} \left\{\exp\left(\hat{\beta}_{Wo} + \widehat{\boldsymbol{\beta}}_{WM} \boldsymbol{M}sub_i\right)/\left(1 + \exp\left(\hat{\beta}_{Wo} + \widehat{\boldsymbol{\beta}}_{WM} \boldsymbol{M}sub_i\right)\right)\right\}^{-1} & for\ Int_i = 1 \\ \left\{1 - \exp\left(\hat{\beta}_{Wo} + \widehat{\boldsymbol{\beta}}_{WM} \boldsymbol{M}sub_i\right)/\left(1 + \exp\left(\hat{\beta}_{Wo} + \widehat{\boldsymbol{\beta}}_{WM} \boldsymbol{M}sub_i\right)\right)\right\}^{-1} & for\ Int_i = 0 \end{cases}$$

We will then use these weights to fit a weighted regression model for the binary outcome, achieving MCID in pain severity at 6-months, applying the same mean model as Step 1 except using the weights w_i. In this approach, the weights in essence "weigh out" the effect of the mediator at 3 months. Comparing the estimates of the intervention effect in the weighted and unweighted models will allow us to estimate how much the intervention effect on the outcome can be explained by each potential mediator. We will further be able to quantify the amount of effect explained by each mediator independent of the other mediators.

17.0. Exploratory (Tertiary) Effectiveness Analyses

We will examine the impact of the active interventions on exploratory outcomes using the same proposed approach for secondary outcomes outlined above in Section 15.0.

18.0. Economic Analyses

We will conduct a full economic evaluation of the CBT-based interventions, compared to usual care, using the framework of cost-effectiveness, including the costs of implementation and maintenance, following best practice in economic evaluation.[26,27] This analysis will be conducted for the Kaiser Permanente clinical sites where the capture of all health care utilization is available through administrative data; inclusion of Essentia Health in these cost analyses is subject to the findings of exploratory analyses related to comprehensiveness of EHR data sources that is being conducted during the UG3 phase. Information on resources used to implement the intervention will come from the trial data collection instruments and from medical office staff, provider interviews, and study staff. We

will include all relevant resources used in the intervention delivery (e.g., training, counseling, fidelity assurance). EHR data will be used to identify and classify health care encounters and prescription medications. Using the framework of cost-effectiveness, we will estimate the incremental cost per additional patient with a MCID in pain severity (30% reduction from baseline), at 12 months, and the Quality-Adjusted Life Year (QALY) gained—utilities will be estimated using the EQ-5D-5L.[11]

Costs to be collected. Medical care utilization and intervention costs will be considered. Medical care utilization includes pharmacy, outpatient visits (including specialty care), inpatient stays, and referrals, and will be costed using standard costing algorithms[12,13] and Medicare fee schedules. In addition to total medical care costs, we will also undertake an analysis of pain-related care focused on utilization linked to pain conditions, identified by diagnostic and procedure codes, and pain-related medications. Intervention costs include program implementation (e.g., training, meetings, and supervision; patient identification, invitation and screening) and delivery (e.g., online hosting, clinician calls). The analysis will take the perspective of the health plan (a principal decision maker for future implementation), so it will include all health system costs of intervention implementation and delivery in clinical settings.

Cost-effectiveness calculations**.** As we have done in prior economic evaluations of trials,[28] we will estimate the cost-effectiveness of the intervention using net benefit regression methods.[29,30] This technique uses a "net benefits" framework, comparing the incremental cost-effectiveness ratio to a range of potential values for a decision maker's willingness-to-pay (WTP) for a unit of health gain. A cost-effectiveness acceptability curve (CEAC) is constructed that illustrates the intervention's probability of being cost-effective at various levels of WTP for a unit of outcome (e.g., cost per QALY of $30,000 to $100,000). The regression framework allows ready evaluation of cost-effectiveness in subgroups (following the intervention's findings). Net benefit regression uses as the dependent variable, net benefit: $nb_i = \lambda \cdot effect_i - cost_i$ (from person-level effect and cost data; λ = WTP level and is varied to construct the CEAC). We will perform sensitivity analyses to assess the applicability of costs to other settings, the estimation of replication costs, and economies of scale.[31]

Health care cost comparisons: We will undertake a comparison of the health care costs between the randomized groups. These comparisons will include overall, and pain-specific costs. The table below describes the proposed categories of health care costs. Health care costs will be analyzed using generalized estimating equations with appropriate link functions.

Categories of Health Care Costs to be Assessed		
	Pain-related[1]	**Total**
Prescription medications		
Ambulatory encounters		
Primary Care		
Physical Therapy, Occupational Therapy and / or Physiatry[2]		
Pain Medicine/Pain Clinic		
Mental/Behavioral Health or Addiction Medicine		
ER or Urgent Care		
Other specialty medical care[3]		
In-patient hospital[4]		

1. Pain-related medications are identified by medication class (i.e., opioids, etc.). Pain-related in-person health care encounters are identified based on ICD-10-CM diagnostic codes.
2. Physical Therapy includes Physical Therapy, Occupational Therapy and Physiatry visits.
3. Specialty medical care includes in-person encounters with any non-primary care department that is not already included in the table.
4. Pain-related hospitalizations have a primary (or principal) pain-related ICD-10-CM diagnostic code.

19.0. Safety Monitoring Analyses

Safety monitoring analyses will be prepared for the external Data Safety and Monitoring Board (DSMB) twice per year to align with the DSMB meeting schedule. **See the Study Protocol and Data and Safety Monitoring Plan for specific details on the data and safety monitoring procedures and reporting guidelines.**

Note that the behavioral interventions being evaluated in this comparative effectiveness trial have been widely used and evaluated in trials with a range of patient types who have similar clinical characteristics to the RESOLVE study population. The trial does not include the use or evaluation of any drug, device or experimental treatment.

19.1. Adverse Events

An adverse event (AE) is any untoward or unfavorable medical occurrence in a human subject, including abnormal signs, symptoms, or diseases, temporarily associated with the subject's participation in the research, whether or not considered related to the subject's participation in the research. Because the RESOLVE study population includes patients with chronic pain, who have fluctuating physical and emotional symptoms as part of the natural course of their condition, such symptoms will not be systematically solicited as part of the trial.

Specifically, AEs will be identified if participants proactively contact the study team at any time to self-report AEs or self-report during a regular study contact; or by intervention or study staff who are interacting with participants regularly throughout the intervention. Since AEs will not be solicited uniformly from all study participants, collection of these events will be biased by study arm (i.e., telephone CBT-CP participants will have greater interaction with study staff by phone and more opportunity to self-report these symptoms and events). AEs will be documented and tracked; a report of the frequency of AEs by site will be reviewed by the PI quarterly and provided to the DSMB on a semi-annual schedule.

Although we considered the possibility of examining other events of special interest particularly those of interest for the HEAL initiative overall (e.g., suicidality), we decided this was not advisable both because there is no clinical or empirical basis for hypothesizing that either of the active interventions would increase risk of events of special interest and because the unequal contact with study staff among participants in the different study arms inherently results in differential opportunities for ascertainment of "events." In this circumstance, between-group comparison of adverse events reported to or discovered by study staff would be biased and hence not meaningfully interpretable or actionable.

19.2. Serious Adverse Events

For this comparative effectiveness trial, a serious adverse event (SAE) is defined as an event that:

- Results in death during a patient's active participation in the trial.
- Results in inpatient hospitalization during a patient's active participation in the trial. (Planned hospitalization scheduled before the enrollment of a study participant is not an SAE.)

Given the relatively large sample size for this trial (N=2,331), we expect some inpatient hospitalizations and deaths to occur among participants. However, none of our previous trials using these types of behavioral interventions have identified study related SAEs.

The following is a summary of the procedures that will be used for monitoring and reporting SAEs; further detail is provided in the DSMP.

Each clinical site will query active participants' electronic health records data every 6 months (in alignment with the DSMB meeting schedule) to identify any deaths or hospitalizations throughout the interval of active participant enrollment in the trial.

For deaths, the date of the event and information needed to facilitate the chart review (described below) will be extracted. For hospitalizations, diagnoses, reason for admission, and length of stay will be extracted to allow the determination of potential relatedness.

A physician within each health care system/clinical site will conduct a chart review for each death to assess its potential relatedness to study procedures and interventions. The findings of each review will be entered into a DSM electronic data capture system hosted by the DCC.

Hospitalizations will not be chart reviewed routinely, however if reports suggest possible relatedness, chart reviews will be conducted on all or a subset of hospitalizations based on DSMB recommendation.

A report summarizing the SAEs and chart reviews will be provided to the DSMB at regular meetings. Any SAEs that are identified during these semi-annual reviews which meet the criteria for immediate reporting to the IRB, NIA PO, and DSMB will be reported according to DSMP procedures.

In addition to the semi-annual EHR queries, SAEs (hospitalizations and deaths) may be identified by participant self-report. If this occurs, the event will be documented and reviewed by the designated physician at the clinical site as part of the semi-annual review described above; it will not be reviewed at the time of reporting.

19.3. Safety Outcomes

The safety outcomes for this study are hospitalizations and deaths (SAEs).

SAEs will be summarized by type and by treatment group in terms of frequency of the event, number of subjects having the event, timing relative to randomization, and relatedness to the study treatment (related, possibly related, not related). Despite the relatively large sample in this trial, we anticipate incidence of possibly treatment related hospitalizations and deaths to be relatively rare so do not propose any formal statistical tests of the SAE data.

19.4. Descriptive Analyses

The safety analyses will be performed using descriptive statistics to quantify hospitalizations and deaths (SAEs) by usual care control arm and separately by each intervention arm. As the study is not powered to test for non-inferiority of the safety events by exposure to the interventions compared to the usual care control, no statistical tests will be performed on group differences in the SAE safety.

19.5. Summary of AEs and SAEs by MedDRA Code and Grouped by Organ System

Because this trial focuses exclusively on behavioral interventions MedDRA codes covering pharmaceuticals, biologics, vaccines, and drug-device combination products are not applicable.

19.6. Safety Monitoring and Recruitment Suspension Rules

The DCC will monitor the safety data and will notify the DSMB chair per the protocol specified in the DSMP. The DSMB chair will provide recommendations regarding the temporary suspension of enrollment and/or administration of study procedures or study interventions and an ad-hoc DSMB meeting will be convened.

20.0. Interim Analyses

There are no planned interim analyses of primary or secondary outcome data before the study is completed. However, if in context of evaluating the safety outcomes the DSMB requests interim effectiveness estimates they will be provided. No formal futility or effectiveness interim analyses will be conducted.

21.0. References

1. Cleeland C, Ryan K. Pain assessment: global use of the Brief Pain Inventory. *Annals of the Academy of Medicine, Singapore*. 1994;23(2):129-138.
2. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *The Clinical journal of pain*. 2004;20(5):309-318.
3. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *The Journal of Pain*. 2004;5(2):133-137.
4. Dworkin RH, Turk DC, McDermott MP, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain*. 12/2009 2009;146(3):238-244. Not in File.
5. Carle AC, Riley W, Hays RD, Cella D. Confirmatory Factor Analysis of the Patient Reported Outcomes Measurement Information System (PROMIS) Adult Domain Framework Using Item Response Theory Scores. *Med Care*. Oct 2015;53(10):894-900. doi:10.1097/mlr.0000000000000413
6. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. Apr 2009;114(1-3):163-73. Not in File. doi:10.1016/j.jad.2008.06.026
7. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 5/22/2006 2006;166(10):1092-1097. Not in File.
8. Full KM, Malhotra A, Crist K, Moran K, Kerr J. Assessing psychometric properties of the PROMIS Sleep Disturbance Scale in older adults in independent-living and continuing care retirement communities. *Sleep Health*. Feb 2019;5(1):18-22. doi:10.1016/j.sleh.2018.09.003
9. Dahlhamer J, Lucas J, Zelaya C, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. *MMWR Morbidity and mortality weekly report*. Sep 14 2018;67(36):1001-1006. doi:10.15585/mmwr.mm6736a2
10. Von Korff M, DeBar LL, Krebs EE, Kerns RD, Deyo RA, Keefe FJ. Graded chronic pain scale revised: mild, bothersome, and high-impact chronic pain. *Pain*. Mar 2020;161(3):651-661. doi:10.1097/j.pain.0000000000001758
11. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. Dec 2011;20(10):1727-36. doi:10.1007/s11136-011-9903-x
12. Hornbrook M, Goodman, M.J., ed. *Adjusting health benefit contributions to reflect risks.* JAI Press; 1991. Hornbrook M, ed. *Risk Based Contributions to Private Health Insurance*.
13. O'Keeffe-Rosetti MC, Hornbrook MC, Fishman PA, et al. A standardized relative resource cost model for medical care: application to cancer control programs. *Journal of the National Cancer Institute Monographs*. 2013;2013(46):106-16. doi:10.1093/jncimonographs/lgt002
14. Mayhew M, DeBar LL, Deyo RA, et al. Development and Assessment of a Crosswalk Between ICD-9-CM and ICD-10-CM to Identify Patients with Common Pain Conditions. *J Pain*. May 23 2019;doi:10.1016/j.jpain.2019.05.006
15. McWilliams LA, Kowal J, Wilson KG. Development and evaluation of short forms of the Pain Catastrophizing Scale and the Pain Self-efficacy Questionnaire. *Eur J Pain*. Oct 2015;19(9):1342-9. doi:10.1002/ejp.665
16. Anderson KO, Dowds BN, Pelletz RE, Edwards WT, Peeters-Asdourian C. Development and initial validation of a scale to measure self-efficacy beliefs in patients with chronic pain. *Pain*. 10/1995 1995;63(1):77-84. Not in File. doi:030439599500021J [pii]
17. DeBar L, Benes L, Bonifay A, et al. Interdisciplinary team-based care for patients with chronic pain on long-term opioid treatment in primary care (PPACT) - Protocol for a pragmatic cluster randomized trial. *Contemp Clin Trials*. Apr 2018;67:91-99. doi:10.1016/j.cct.2018.02.015
18. Peterson K, Anderson J, Bourne D, Mackey K, Helfand M. Effectiveness of Models Used to Deliver Multimodal Care for Chronic Musculoskeletal Pain: a Rapid Evidence Review. *J Gen Intern Med*. May 2018;33(Suppl 1):71-81. doi:10.1007/s11606-018-4328-7

19. Wang M, Fitzmaurice GM. A simple imputation method for longitudinal studies with non-ignorable non-response. *Biometrical Journal* 2006;48(2):302-318.
20. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. Apr 1 2004;159(7):702-6.
21. Liang KY, Zeger SL. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*. 1986;73(1):13-22.
22. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. Mar 1986;42(1):121-30.
23. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. Dec 1988;44(4):1049-60.
24. RM B, DA K. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 12/1986 1986;51(6):1173-1182. Not in File.
25. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. *J Am Stat Assoc*. 2018;113(523):1112-1121. doi:10.1080/01621459.2017.1305274
26. *COST-EFFECTIVENESS in HEALTH and MEDICINE*. Oxford University Press, Inc.; 1996:425.
27. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *Jama*. Sep 13 2016;316(10):1093-103. doi:10.1001/jama.2016.12195
28. Smith DH, O'Keeffe-Rosetti M, Owen-Smith AA, et al. Improving adherence to cardiovascular therapies: an economic evaluation of a randomized pragmatic trial. *Value in Health*. 2016;19(2):176-184. NOT IN FILE.
29. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health economics*. Jul 2002;11(5):415-30. doi:10.1002/hec.678
30. Hoch JS, Rockx MA, Krahn AD. Using the net benefit regression framework to construct cost-effectiveness acceptability curves: an example using data from a trial of external loop recorders versus Holter monitoring for ambulatory monitoring of "community acquired" syncope. Comparative Study
Research Support, Non-U.S. Gov't. *BMC Health Serv Res*. 2006;6:68. doi:10.1186/1472-6963-6-68
31. Meenan RT, Stevens VJ, Hornbrook MC, et al. Cost-effectiveness of a hospital-based smoking cessation intervention. Clinical Trial
Controlled Clinical Trial
Research Support, U.S. Gov't, P.H.S. *Medical care*. May 1998;36(5):670-8.