Study Protocol and Statistical Analysis Plan

Title: Digital Clinical Hypnosis for Chronic Pain Management

NCT number: NCT06050083

Document date: 08/21/2025

Study Protocol

Overview of Study 1 and Study 2

This protocol describes two studies that are a part of this project. They use the same experimental design to evaluate the feasibility of the digital program. The primary difference is that first study (Study 1) recruits individuals with chronic low back pain and the second (Study 2) recruits individuals with general chronic pain (which includes individuals with chronic low back pain as well as individuals with chronic pain as a secondary condition to another health problem, such as spinal cord injury). Both studies aim to enroll 50 participants, which, given an estimated 20% dropout rate, would allow for complete data from at least 40 participants in each study.

Both studies use a 2-arm prospective randomized controlled trial design to evaluate the feasibility of the program (called *Rose*) over an 8-week period. Participants are randomly assigned to one of two groups: *Rose* access for a full 8 weeks (Access group) or *Rose* access for 4 weeks (Waitlist control group). The Waitlist group will not have access to *Rose* for the first 4 weeks of the study and then have access to *Rose* for the final 4 weeks. The 4-week assessment point is the primary study endpoint. Primary (feasibility) and secondary (clinical) outcomes using the *Rose* program and self-report measures using the Research Electronic Data Capture (REDCap) program, hosted by the University of Washington Institute of Translational Health Sciences (ITHS). Participants complete the self-report measures at three timepoints: baseline (before randomization), 4 weeks after randomization (primary endpoint), and 8 weeks after randomization.

Measures

Feasibility. The studies has five primary feasibility outcome domains: (1) frequency of listening to a hypnosis session; (2) participant satisfaction with *Rose*; (3) interest in having continued access to *Rose* after study completion; (4) willingness to pay for that continued access; and (5) system usability. For each feasibility outcome, we have *a priori* specific criteria for determining if *Rose* was feasible.

Frequency of access. The *Rose* application tracks how frequently participants listened to a session. In this domain, we propose that *Rose* would be feasible if at least 50% of the Access participants who were offered access to *Rose* used it at least once/week. This criterion was selected as clinical hypnosis sessions are usually scheduled to occur once/week, making this frequency a standard for this treatment.

Satisfaction with *Rose*. We assess satisfaction with *Rose* at the 4-week timepoint (i.e. 4-weeks after randomization) using a 5-point global treatment satisfaction scale ("Taking all things into account, how satisfied are you with using the Rose web application?"), whose response options ranged from 1 ("Very satisfied") to 5 ("Very dissatisfied"), or "Prefer not to Answer." We determined that *Rose* would be feasible if ≥ 50% of the participants who listened to at least one *Rose* session would rate being either "Somewhat satisfied" or "Very satisfied" with *Rose*.

Interest in continued Rose access. We assess interest in continued access to *Rose* by asking, that if the program were available, "... how interested would you be in downloading and using this application?" Participants respond to this question using a 5-point Likert scale ranging from 0 = "I would not be interested" to 5 = "I would be extremely interested." We determined that *Rose* would be feasible if ≥ 50% of the

Access group participants who used *Rose* per protocol (i.e., at least 4 times during the four weeks after randomization) rated themselves as being at least "... a little interested" in using Rose.

Willingness to pay for access to *Rose*. We also assess participants' willingness to pay for continued *Rose* access after finishing the study. However, after the study began, we discovered that the question had been phrased incorrectly. The error was corrected when the error was discovered, but most of the Study 1 participants and four of the Study 2 had responded to the incorrectly worded question when the question was corrected. We therefore present the findings for this question only for the Study 2 participants in the Study 2 participants in the Access group who used *Rose* per protocol. We determined *a priori* that *Rose* would be feasible if ≥ 40% of the Access group participants who used *Rose* per protocol endorsed a willingness to pay for continued access.

Usability. Finally, we assess the usability of *Rose* using the System Usability Scale (SUS)(Lewis, 2018), a common 10-item measure of the overall usability of applications (sample item, "I thought the product was easy to use"). The cutoff we selected as the minimum average needed to determine that *Rose* was feasible was a score of ≥ 68, which is the average SUS score across a large variety of software products (i.e., if Rose's SUS score is ≥ 68, it has an above average level of usability; (Banger, Kortum, & Miller, 2008)).

Secondary outcomes: Changes in clinical domains. Study 1 and Study 2 were both feasibility studies and are therefore not powered to detect statistically significant effects on any clinical outcome. However, the primary purpose of *Rose* is to improve pain-related quality of life domains for individuals who access and listen to *Rose's* sessions, and these two feasibility studies were designed to provide data which could be used to inform a full clinical trial, if *Rose* is found to be feasible. A secondary aim of the two feasibility studies was therefore to provide preliminary estimates of the effects of *Rose* on a number of pain-related clinical outcomes to: (1) determine if anticipated improvements in these outcomes could be observed and, if so, to help (2) identify (a) the outcome domains that should be assessed in a fully-powered clinical trial, and (b) the measures that should be used to assess those domains in such a clinical trial.

We assess six clinical outcome domains: (1) immediate (i.e., pre- to post-session) changes in pain intensity; (2) change in average pain intensity from baseline (i.e., at randomization) to four weeks after randomization (primary endpoint) and to eight weeks after randomization (secondary endpoint); (3) change in pain interference from baseline to 4-weeks and from baseline to 8-weeks; (4) change in anxiety from baseline to 4-weeks and baseline to 8-weeks; (5) change in sleep disturbance from baseline to 4-weeks and baseline to 8-weeks; and (6) change in opioid use from baseline to 4-weeks and baseline to 8-weeks after randomization.

We measure pain intensity for the first two clinical outcomes using a 0-10 Numerical Rating Scales (NRS) of pain intensity. The endpoints of the NRSs used here were 0 = "No pain" and 10 = "Pain as bad as you can imagine." NRSs have a great deal of evidence supporting their reliability and validity as measures of pain intensity (Jensen, 2018) and are recommended over other ratings scale in pain clinical trials (Dworkin et al., 2005). To assess immediate changes in pain intensity associated with

listening to a *Rose* session, the *Rose* application prompts participants with a screen/question asking participants to rate their *current pain* on the 0-10 NRS just before and just after each listening session that included hypnotic suggestions for comfort. To assess change in average pain intensity from baseline to the 4-week and the 8-week assessment points, participants are asked to rate their "Average pain intensity in the past week" with the NRS at baseline (immediately after randomization), and then 4- and 8-weeks after randomization.

Pain interference, anxiety, and sleep disturbance at baseline, week 4, and week 8 are assessed by administering the Patient Reported Outcomes Measurement Information System (PROMIS) 8-item short form scales assessing these domains (PROMIS, 2024). PROMIS measures ask respondents to consider the magnitude or frequency of each symptom in the past seven days listed on the short forms using 5-point Likert scales with the response options responses varying as a function of the item (e.g., for magnitude, "Not at all" to "Very much"; for frequency, "Never" to "Always"). Responses are summed and then transformed to T-scores with a mean of 50 and a SD of 10 in the normative samples. All PROMIS measures are scored such that higher scores indicate higher levels of the domain being assessed (i.e., here, more pain interference, anxiety, or sleep disturbance). Evidence supports the validity and reliability of each of these measures in a variety of clinical populations, including individuals with chronic pain (Bhandari et al., 2022; Chimenti et al., 2021). The baseline internal consistencies (Cronbach's alphas) for these measures were: .95, .95, and .90 for Study 1 and .93, .86. .91 for Study 2 the PROMIS Pain Interference (8-item Short Form), PROMIS Anxiety (8-item Short Form), and PROMIS Sleep Disturbance (8-item Short Form) scales, respectively.

Opioid use is assessed by a self-report question asking participants if they are taking any opioid medications (with opioid examples listed, such as oxycodone, hydrocodone, and fentanyl). If the participant responded "Yes," a sub-question appeared asking for: medication name, dosage, and frequency of use in the past seven days. These participant responses were then converted to Morphine Milligram Equivalents (MME), which refers to “the amount of morphine that an opioid dose is equal to when [the opioid is] prescribed.” MME converts different opioid medication dosages into a single unit and is calculated by comparing the results from two publicly available MME calculators (Hudson et al., 2018).

Procedures

Consent and enrollment

Study 1 (Chronic Low Back Pain). The Study 1 participants are recruited using Research Match, a national research participant registry managed by the National Institute of Health. We use a multi-step recruitment process to identify eligible participants and focus on recruiting a racially diverse sample. First, we set the Research Match location filters to be anywhere in the United States, health condition filters to '"Chronic Low Back Pain," demographic filters to any sex/gender, and race/ethnicity filters to all non-White or Hispanic/Latinx individuals. After screening and enrolling participants from this first recruitment wave, we evaluate how close we were to our recruitment goals for sex and race. Our goal is to have racial diversity that is at or above average racial diversity percentages for the Seattle metropolitan area. That is: ≤ 47%

White Non-Latino/Latina, ≥ 11% Asian, ≥ 6% Black or African American, and ≥10% More than one Race, and self-identified ethnicity (goal: ≥ 6% Hispanic or Latinx). Next, we adjust the Research Match filters to include White/Caucasian, Non-Hispanic Latinx adults until we reached our recruitment goal of N = 50. Potentially eligible participants received a study email invitation from Research Match with a link to our online REDCap screening form. The online screening form provided a UW Institutional Review Board (IRB) approved study summary and collected participants' contact information and screening question responses.

The Study 1 screening form assesses for all study inclusion and exclusion criteria except for exclusion criteria (1), (2), and (3), which are assessed via telephone interview later. The inclusion criteria include: (1) 18 years old or older; (2) self-reported chronic low back pain (i.e., pain in the low back that has been ongoing for 3 months or more; and occurs on at least half the days in the past 3 months); (3) average pain intensity in the past week of ≥ 4 on a 0-10 scale, where 0 = “No Pain” and 10 = “Pain as bad as you can imagine”; (4) reads, speaks, and understands English; (5) regular access to the internet per self-report; and (6) access to a device (smartphone, desktop, laptop, tablet, etc.) that can connect to the internet. The study exclusion criteria include: (1) history or current diagnosis of primary psychotic or major thought disorder (i.e. Schizophrenia, Psychosis, Bipolar Disorder) with current active symptoms such as active delusional or psychotic thinking (self-report) within the past 5 years per self-report; (2) psychiatric hospitalization within the past 6 months per self-report; (3) psychiatric or behavioral conditions, including active suicidal ideation or intent to harm oneself or others, in which symptoms were unstable or severe within the past 6 months as disclosed during self-report screening; (4) active malignancy (e.g., cancer not in remission) or actively in cancer treatment per self-report; (5) planned surgery in the next six months that is related to the back pain per self-report; and (6) currently receiving or have received hypnosis treatment for any pain condition per self-report.

Potential participants who met the inclusion and exclusion criteria based on the REDCap screening form are contacted by study staff via telephone within 2 days of staff receiving a REDCap notification of a completed online screening form. During the phone call, study staff administer questions to assess the potential participant’s history or current diagnosis of primary psychotic or major thought disorder, whether they had been hospitalized for a psychiatric treatment within the past 6 months, and the presence of active suicidal ideation or intent to harm oneself or others. Participants that endorse a history of being diagnosed with schizophrenia, psychosis or bipolar disorder but did not have any active symptoms were still eligible. Participants that endorse active suicidal ideation or any of the above nine follow up questions regarding symptoms that could interfere with participation (e.g., delusional thinking) were referred to the study PI (MPJ), who is a licensed clinical psychologist, for a final eligibility decision (and potential referral to clinical resources).

Study 2 (any type of chronic pain). The recruitment process for Study 2 isalmost identical to Study 1. The only difference is that instead of requiring that the participant have chronic low back pain, the Study 2 participants need to endorse having any type of chronic pain.

Informed Consent. For both Study 1 and Study 2, participants provide verbal informed consent over the telephone with study staff. The UW IRB approved a waiver of written informed consent for this study.

Statistical Analysis Plan

Descriptive analyses of *Rose* use. We will compute the total number and percent of times sessions of each different duration were listened to (i.e., 5, 10, or 20 minutes) and the number of percent that each different voice were listened to across the entire study period for both groups.

Feasibility analyses. To evaluate the feasibility of *Rose*, and for both studies, we will first compute the mean, standard deviation, median and range for the number of times participants in the Access conditions listened to sessions as well as the number and percent of the participants who listened to a session at least four times during the first four weeks of access (i.e., per protocol). To evaluate satisfaction with use of *Rose*, we will compute the average satisfaction rating for those participants in the Access groups who listened to at least one session during the first four weeks of access, as well as the number and percent of the responses to each of the satisfaction response options. To evaluate interest in continued access to *Rose*, we will compute the average of the interest ratings of the participants in the Access groups who used *Rose* per protocol (i.e., at least four times during from baseline to the 4-week assessment point), as well as the number and percent of the responses to each of the interest question response options. To understand the participants' willingness to pay for continued access to *Rose*, we will compute the number and percent of responses to each of the payment question response options for the Access group participants in each study who had used *Rose* per protocol. Given the inherent differences between the two studies, Fisher's exact tests will be used to compare satisfaction and interest for continued access between the two studies. To assess the usability of *Rose*, we will compute the average SUS for participants in the Access groups of each study who had used at least once.

Clinical effects analyses. In order to estimate the possible effects of *Rose* on clinical outcomes, we will conduct two types of analyses. First, to estimate the immediate effects of the *Rose* sessions that target pain intensity, we will compute the ranges and average pre- to post-session changes in current pain intensity ratings for each participant individually (including participants in the Access condition who had access to 8 full weeks from baseline to the 8-week assessment point, and participants in the waitlist condition who had access for 4 weeks from the end of week 4 to the end of week 8 assessment point) and compute the median, minimum, and maximum of these values. Median statistics will be computed as the measure of central tendency given that the distributions of these changes were anticipated to be positively skewed (i.e., we anticipate that most participants would access *Rose* about four times during every 4-week period of access, but a fewer number might access *Rose* many times). Of interest is the general consistency of responses to the *Rose* sessions both within and between individuals. We will also conduct a linear mixed-effects model for the pre- to post-session change in pain intensity, including a random intercept for each participant to account for clustering (one for each study), to estimate the effect size and

significance levels of these changes. Given the possibility that participants who found *Rose* to be particularly effective may choose to listen to more sessions than those who found *Rose* to be less effective, as a sensitivity analysis, we will assess if there is evidence of informed cluster sizes by testing if the effectiveness of *Rose* is associated with the number of times sessions were listened to. This allows us to determine if the magnitude of the pre- to post-session improvements in pain intensity and mood are associated with frequency of listening (i.e., if those who reported more benefits tended to listen more often, which could bias the estimates of average benefit).

Next, we will calculat the mean and standard deviation for each of the five continuous outcomes and baseline, 4-weeks, and 8-weeks stratified by treatment condition (Access vs. Waitlist group). Standardized effects will be calculated to determine the within group change between baseline and both 4- and 8-weeks as well the between group difference in these changes. As this is feasibility study, no rigorous test of significance will be performed. However, 95% confidence intervals will be computed for the between group effect sizes to determine the range plausible effect sizes in the population.

References cited

Banger, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluationof the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574-594.

Bhandari, R. P., Harrison, L. E., Richardson, P. A., Goya Arce, A. B., You, D. S., Rajagopalan, A., . . . Sil, S. (2022). Clinical Utility of CAT Administered PROMIS Measures to Track Change for Pediatric Chronic Pain. *J Pain, 23*(1), 55-64. doi:10.1016/j.jpain.2021.06.009

Chimenti, R. L., Rakel, B. A., Dailey, D. L., Vance, C. G. T., Zimmerman, M. B., Geasland, K. M., . . . Sluka, K. A. (2021). Test-Retest Reliability and Responsiveness of PROMIS Sleep Short Forms Within an RCT in Women With Fibromyalgia. *Front Pain Res (Lausanne), 2*, 682072. doi:10.3389/fpain.2021.682072

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., . . . Immpact. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain, 113*(1-2), 9-19. doi:10.1016/j.pain.2004.09.012

Hudson, T. J., Painter, J. T., Gressler, L. E., Lu, L., Williams, J. S., Booth, B. M., . . . Edlund, M. J. (2018). Factors Associated with Opioid Initiation in OEF/OIF/OND Veterans with Traumatic Brain Injury. *Pain Med, 19*(4), 774-787. doi:10.1093/pm/pnx208

Jensen, M. P. (2018). Measurement of pain. In S. M. Fishman, J. C. Ballantyne, & J. P. Rathmall (Eds.), *Bonica's management of pain*. Media, PA: Williams & Wilkins.

Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human-Computer Interaction, 34*, 577-590.

PROMIS. (2024, 2024). The NIH Patient-Reported Outcomes Measurement Information System. . Retrieved from http://www.healthmeasures.net/explore-measurement-systems/promis