University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Download date: 04. May. 2023

A prospective, double-blind, pilot randomized controlled trial of an 'embodied' virtual reality intervention for chronic low back pain in adults

Christopher Eccleston[1,2], Emma Fisher[1,3], Sammeli Liikkanen[4], Toni Sarapohja[4], Carina Stenfors[4], Satu K. Jääskeläinen[5], Andrew S.C Rice[6], Leena Mattila[4], Taru Blom[4], J. Raymond Bratty[4]

Institutions

1. Centre for Pain Research, University of Bath, Bath, UK.

2. Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium.

3. Cochrane Pain, Palliative, and Supportive Care Review Groups, Oxford University Hospitals, Oxford, UK

4. Orion Corporation Orion Pharma, R&D, Espoo Finland.

5. Department of clinical neurophysiology, Turku University Hospital and University of Turku, Turku, Finland.

6. Pain Research, Department of Surgery and Cancer, Faculty of Medicine, Imperial College, London SW10 9NH, UK.

Corresponding author: Christopher Eccleston, Centre for Pain Research, University of Bath, Claverton Down, Bath, BA27AY. [c.eccleston@bath.ac.uk]

Abstract

Adults with chronic low back pain, disability, moderate to severe pain, and high fear of movement and re-injury were recruited into a trial of a novel, automated, digital therapeutics virtual reality psychological intervention for pain (DTxP). We conducted a three-arm, prospective, double-blind, pilot randomized controlled trial comparing DTxP with a sham placebo comparator, and an open standard care. Participants were enrolled in for 6-8 weeks, after which, the standard care control arm were re-randomized to receive either the DTxP or sham placebo. Forty-two participants completed assessments at baseline, immediately post-treatment (6-8 weeks), 9-week, and 5-month follow-up. We found participants in the DTxP group reported greater reductions in fear of movement and better global impression of change when compared with sham placebo and standard care post-treatment. No other group differences were noted at post-treatment or follow-up. When compared to baseline, participants in the DTxP group reported lower disability at 5-month follow-up, lower pain interference and fear of movement post-treatment and follow-up, and lower pain intensity at post-treatment. The sham placebo group also reported lower disability and fear of movement at 5-month follow-up compared to baseline. Standard care did not report any significant changes. There were a number of adverse events, with one participant reporting a serious adverse event in the sham placebo which was not related to treatment. No substantial changes in medications were noted, and participants in the DTxP group reported positive gaming experiences.

Summary

A novel, automated, virtual reality psychological intervention for people with low back pain reduced fear of movement compared to sham placebo and standard care control.

1. Introduction

Fear of movement, characterized by the dominant belief that pain and re-injury are the likely consequence of movement, can maintain and exacerbate disability associated with chronic pain [6; 11]. The rigidity of beliefs of inevitable harm, distracting worry and self-defeating cognition, and subsequent social withdrawal, are all targets of intervention in evidence supported cognitive behavioral therapy (CBT) and CBT informed rehabilitation [20; 42; 66; 67].

Access to CBT informed rehabilitation, however, is rare, in part due to a shortage of health care professionals with the appropriate skill and experience, and reliance on delivery methods requiring the physical co-location of patients and health care professionals. Internet and telemedicine solutions have gone some way to allow for remote treatment delivery [19; 53], and telemedicine adoption has recently been accelerated by the public health measures introduced in response to the COVID-19 pandemic [15]. Despite growth in experience with technologically mediated consultation, widespread adoption of digital therapeutics is still some way off.

Virtual Reality (VR) technologies offer significant opportunities in psychotherapy generally [39] and in pain management and rehabilitation specifically.[34; 63] VR can enable existing treatments to be offered at scale through automation, but it can also improve the content and impact of both assessment and intervention. Although there are many reports of VR for analgesia [30; 49; 56] and sedation [54], there are few examples of attempted use in chronic pain.[34] Notable exceptions include RCTs of skills-based interventions with a focus on emotional self-regulation [4; 12; 24; 25] and rehabilitation interventions with a focus on education and increasing movement.[23; 61]

We designed, developed, and tested a VR intervention for use with adults with chronic pain who have high disability, and high fear of movement and re-injury. The content was maximized to make use of the core properties of virtual reality: immersion, exploration, and experimentation. The intervention requires daily goal-oriented movement that increases in both intensity and range of movement over time [60]. The content is guided by an 'embodied' model of the psychology of pain in which disability and distress arise from a) the privileged access to consciousness of threat-relevant interoception (i.e., bodily sensations are more likely to be attended to, interpreted as threatening, and acted upon), b) avoidance behavior maintained with reinforcement by immediate and general safety signals, and generalized to multiple contexts, and c) consequent social and cognitive failure supported by self-defeating behavior and cognition [13; 14; 59]. Moving without concern for harm, experiencing positive outcomes after self-determined action, and safely experimenting with core beliefs can all be delivered in an immersive environment within VR. Further, given the public health goal to scale access to evidence-based treatments, the intervention was protocolized and automated, so it could run without access to an interpreting health care professional.

We report a pilot randomized controlled trial comparing the active VR intervention with a sham placebo comparator and a standard care group. The primary aim of this study was to determine any comparative efficacy and report any adverse events. We also report on experience with the VR interventions.

2. Methods

2.1 Trial design

A pilot three-arm prospective, double-blind, randomized controlled trial comparing a digital therapeutics software solution for chronic pain known here as

Digital Therapeutics for Pain (DTxP), a sham placebo comparator, and an open standard care. Participants followed their allocated treatment for 6-8 weeks and were assessed immediately post-treatment and again 9 weeks after randomization to ensure a fixed assessment timepoint for all participants. After unblinding, participants continued in an open label extended follow-up period for a further 3 months (end of study-5 months post-randomization). The treatment was protocolized, automated, and delivered by a machine mentor which was blind to allocation. Further, participants and some study staff were blind to allocation in the DTxP and sham placebo arms. Participants in the standard care arm were open to allocation but were re-randomized to receive the DTxP or sham placebo after the week 9 follow-up, and then blinded to their new allocation (Figure 1). The study was pre-registered on clinicaltrials.gov (NCT04225884) and the trial design and statistical protocols are available in the supplementary material. Specifically, the protocol outlines initial proof of concept with 30 participants and potential expansion arms for up to 100 participants, which could be based on learnings from this study had the device failed, or for patients with different aetiologias of pain. The expansion arms were not utilized after results of the feasibility study were reviewed. The Helsinki University Ethical Committee gave ethical approval for the study (approval number HUS/3111/2019).

2.2 Participants and setting

The inclusion criteria were: 1) adults, $\geq$18 years of age, who 2) reported low back pain for at least 3 months, with average pain intensity of $\geq$4/10 over the past week on a 0-10 numeric rating scale, Oswestry Disability Index (ODI) $\geq$26%, and medium (34-41) or high (42-68) Tampa Scale of Kinesiophobia (TSK) score; 3) provided written informed consent; 4) were Finnish speaking; 5) had a clear, flat

ground surface at home of at least 2 meter squared and could interact with a digital therapeutics intervention for pain; and 6) could bend without severe pain.

Several exclusion criteria were applied including: 1) a history of epilepsy, migraine, vertigo or psychosis; 2) a confirmed diagnosis of cancer; 3) susceptibility to motion sickness requiring treatment; 4) pregnancy; 5) current physiotherapy that contraindicated the goals of the DTxP; 6) severe or acute structural pathologies that the intervention could make worse, as assessed by a physician, 7) having had psychotherapy in the previous 2 years, or currently receiving psychotherapy (so to avoid any interactions of the experimental conditions in this trial with previous therapies, and to protect blind allocation), 8) reported any condition that affected posture or balance; 9) any prior participation in a digital therapeutics intervention for pain study (e.g., the feasibility testing study) or had participated in any part of the current study.

We recruited participants from January 2020 to October 2020 through internet advertisements (via www.orion.fi) and social media (via Facebook). Participants were screened using the above criteria and then randomized into the arms of the treatment.

2.3 Procedures

2.3.1 Randomization and blinding Randomization was performed using a random numbers generator by an independent randomization expert who was not part of the study team. Participants were randomized to a group using a 1:1:1 allocation ration and block size of 3, stratified by their Tampa Scale of Kinesiophobia (TSK) score (medium or high score). The randomization expert then sent the randomization list via email to the unblinded study members who allocated the participants into the trial condition following the randomization order. For the second randomization of the standard care group into the DTxP or sham placebo group after the 9-week

assessment, a 1:1 allocation ratio and block size of 2 was used, stratified by medium or high score on the TSK. Further help and instructions were given by an unblinded study member, without a blinded study team member being present.

Participants were not informed of their treatment allocation and treatments were only referred to by either 'treatment A' or 'treatment B' on study materials. Participants were not allowed to compare treatments. Two groups of personnel were formed: an unblinded study team and a blinded study team.

The unblinded study team's responsibilities included:

- o Not revealing treatment allocations to anyone else,
- o Issuing the correct, allocated device to the subject and supervising the training,
- o Technically supporting participants throughout the duration of the study,
- o Calling the subject to elicit adverse events, device deficiencies or adverse device effect, recording this and then entering the information to the electronic shared database in such a way that the treatment allocation remained concealed.

The blinded study team group included the PI and staff conducting the analyses.

2.3.2 Study flow - Participants were screened and then gave consent to be involved in the study. Participants attended the Testing Unit during the first session. During this session, participants were given a VR wireless headset and hand controllers, corresponding to their randomization number. Participants were shown how to use the headset, hand-held controllers and all movements required by the treatment. The DTxP or sham program was pre-loaded. Early during recruitment, in-person sessions were suspended due to COVID-19 and replaced with training videos, instructions, and phone support. VR headsets were delivered to participants' homes.

During home-use, participants were instructed to use the VR technology on a clear, flat ground surface 2m x 2m in size to prevent injury. Participants range of movement and stability were assessed by staff at the first session. Participants completed assessments before treatment (baseline), post-treatment (6-8 weeks), 9-week follow-up and end of study (5 months post-randomization). Some outcomes were assessed more frequently as outlined in our protocol but not reported here. After 9-week follow-up assessment, participants in the standard care group were re-randomized into the DTxP or sham placebo comparator for 6-8 weeks and completed follow-up at 9-weeks and 5-months post-randomization.

2.4 Treatment groups

2.4.1 Design Model. We adopted a 'design thinking' user-centric methodology to create both the DTxP and sham placebo intervention [36] which involves the iterative use of abduction to develop solution focused designs. Multiple stakeholders, were interviewed for their insights and input at the design, development, and implementation stages, including patients, pain practitioners, pain research experts, UX design experts, and designers. Multiple bespoke design tools were created, to track content fidelity, quality control and safety (available on request). The DTxP and sham placebo intervention were principally designed by Professor Eccleston, who has a UK Health Professionals Council registration to practice health psychology (PYL03729), 15 years' experience in interdisciplinary residential chronic pain management, and has written on the methods of embodied psychology. The protocols, and inclusion and exclusion criteria were reviewed by Professors Rice and Jääskeläinen, and the movements promoted in the activity were reviewed by experts from the Finnish Spinal Health Association

2.4.2 Hardware and Software. Both the DTxP and sham placebo were built and tested using Unity v2019.3.7f1 and delivered on an off-the-shelf OCULUS Quest and Touch VR headset and accompanying Touch hand-held controllers given to participants for at home use. All the scenes in VR run at a refresh rate minimum of 70Hz.

2.4.3 Duration and exposure. Both the DTxP and sham placebo were designed to last 15-60 minutes per session. There were five sessions scheduled for each week, resulting in 30 unique days that were designed to be delivered over 6-8 weeks.

2.4.4 Virtual Environments. Both treatments included an 'inside space' rendered as a summer cabin and an 'outside space' rendered as a lakeshore and copse with fruit trees. Ambient sound was present in the outside space commensurate with summer wildlife.

2.4.5 Digital therapeutics software solution for chronic pain (DTxP) – Participants completed a fully immersive virtual reality experience based on a cognitive behavioral intervention maximized to make use of the core properties of virtual reality: immersion, exploration, and experimentation. The intervention is guided by an 'embodied' model of the psychology of pain aimed to reduce fear of movement and re-injury by creating a safe virtual environment to encourage exploration and repeated movement.

Participants always enter the virtual world via the inside space (figure 2). When in the cabin they encounter a virtual (disembodied) mentor. The mentor communicated with a male voice with optional text script that appeared in view for participants to read. The mentor acted as a guide. He welcomed participants, helped them navigate the space, gave instructions on tasks, presented the psychological content inviting reflection, and set homework in the real (non-VR) world. The mentor invited the

participant to the outside space where they engaged in a gamified 'fruit-picking' activity (figure 3). The task involved picking or stacking geometrically shaped fruits. Fruits from trees appeared in all four quadrants of their peripersonal space. Their location and frequency increased through the experience. During the stacking task, the fruits needed to be stacked in a particular order so the shapes could fit together. The activity was designed to promote fine and gross motor movements typically avoided by low back pain patients fearful of movement and re-injury.

The behavior change content was divided into 24 modules. In each five-day period new material was delivered on the 1st, 3rd, and 5th day. The intervening two days (2nd and 4th) were practice days in which no new material was delivered, but participants could revisit material. Day five always involved homework-setting (excepting the final day) and day one always involved homework review (excepting the first day). Each five-day period had a general theme: orientation, education and exploration, problem solving, setback planning, use of peripersonal space, or consolidation and generalization.

The behavior change content was delivered by the mentor in the inside space before going outside. The material was organized around four behavior change themes. First, content was aimed at building and maintaining a *working alliance* between mentor and participant. In particular, the mentor always presented a rationale, offered education, and adopted a relational stance with compassionate regard. Where possible, self-deprecating humor was employed. Specific material included education on the relationships between pain and tissue damage, goal and value determination, reflection on goal-value discrepancy, goal setting and pacing, and positive reinforcement for engagement and activity. Second, the intervention had a strong focus on *embodied reactivity*; participants were encouraged to move within

their peripersonal space, owning and being agents within that space, and engaging in a full range of motion physical activities, repeatedly. The mentor explained the rationale for the activity in the outside space, but also encouraged and rewarded engagement. Third, we promoted *courageous engagement* to counter feared movement and activity by providing education on avoidance behavior, promoting behavioral experiment in confronting feared activity, suggesting reflection on movement achieved despite beliefs of inability or fear of injury and pain, and providing reinforcement for engagement. Fourth, there was material promoting *mastery*, with insight into common cognitive and social difficulties, and techniques offered to promote problem solving, reflection on change, and increased self-efficacy for change.

More details of the content of the DTxP are available as appendices, including Appendix A showing the experience flow over the 30 days, Appendix B showing the themes, module placement, module labels and content, and Appendix C shows the four behavior change themes, description of content, and how the modules map onto the themes. For comprehensiveness, we completed the TiDieR checklist on intervention description in Appendix D.[31]

Three visual metaphors were used in the inside space to support interaction. First, a map of the virtual world served as a device for participants to teleport to the outside space (a book next to an umbrella in the shore area of the outside space allowed participants to teleport back inside). Second, after each module was completed a module book appeared on a bookshelf which allowed participants to see progress and to re-run material if wanted. And third, a growing plant was used to show progress in the outside 'fruit-picking' activity. The mentor guided participants to see changes in these objects and verbally rewarded progress.

2.4.6 Sham placebo comparator – Participants in the sham placebo comparator received the same VR headset as the participants in the DTxP arm. In line with the VR-Core trial guidelines we selected the most stringent comparator possible, of 3-D *'…immersive and active experiences within a headset.*" [5]. The virtual environment showed similar environments to the DTxP arm and could navigate both inside and outside. None of the content provided in the DTxP was presented. Instead, the mentor invited participants to relax and enjoy the same environments as used in the DTxP. The same mentor communicated with a male voice with optional text script that appeared in view for participants to read. The mentor only welcomed the participant and invited them to stay and enjoy the spaces, and then said goodbye, inviting them to join the next day.

2.4.7 Standard care – Participants in the standard care condition received no intervention, no instructions or education. During the treatment phase, participants continued with their standard care as usual for 6-8 weeks. Participants visited the study site for screening on day one and assessment at 6-8 weeks after randomization. For participants enrolled during the COVID pandemic and therefore could not attend an in-person visit, procedures were undertaken by video call. At the end of the first treatment phase, participants were re-screened for inclusion and completed questionnaires, before being re-randomized to either the DTxP or sham placebo comparator.

2.4.8 *Translation* – All material was written in English and then translated into Finnish, and then independently back translated into English to check for meaning. Any corrections were introduced to the final Finnish version.

2.5 Measures

All measures were delivered in Finnish. The primary endpoint was post-treatment (6-8 weeks). Disability, fear of movement, numerical rating scale (NRS) for pain intensity, and patient clinical global impression of change measures were assessed post-treatment, at 9 weeks, and at end of study (5-month follow-up). To reduce burden on participants, other measures were only assessed at post-treatment and 9-week follow-up.

2.5.1 Disability – Participants completed the Oswestry Disability Index (ODI version 2, [2; 46], a self-report measure assessing physical disability in participants with chronic pain. The ODI includes 10 sections; participants choose one statement from each section which was rated from 0-5, and then summed. Scores ranged from 0-50. We dichotomized responses; participants reporting ≤22 were categorized as responders and participants >23 as non-responders [64].

2.5.2 Pain interference - Participants completed the Patient Reported Outcomes Information System (PROMIS) 6-item Pain Interference Score (6b [1], which ranges from 6-30 points with higher scores indicating higher levels of pain interference. Participants responded to items on a scale from 1-5 (1 = not at all, 5 = very much). This measure was assessed at baseline and post-treatment. The raw score from the questionnaire was then translated into a t-score to provide a standardized score with a mean of 50 (SD=10). Please note, the conversion to a t-score was a post-hoc decision as is best practice with this measure.[1]

2.5.3 Fear of movement and re-injury – Participants completed the Tampa Scale of Kinesiophobia [17; 35; 41], which includes 17 items assessing beliefs about pain-related movement and possible further pain and reinjury using a 4-point Likert scale from strongly agree to strongly disagree. Higher scores indicate higher fear of movement and re-injury.

2.5.4 Pain intensity – Participants completed a Numerical Rating Scale [48] at baseline, post-treatment, and end of study to assess the average pain, maximum pain, and minimum pain over the last week. Participants responded on a 0-10 scale (0=no pain at all; 10=worse pain imaginable). Participants also completed the PROMIS pain intensity 3a measure to assess current, average, and worst pain [7] at baseline and post-treatment. Similar to the PROMIS 6b assessment, scores were translated to a t-score to provide a standardized mean of 50 (SD=10). Please note, the conversion to a t-score was not identified in the protocol but is best practice with this measure.

2.5.5 Pain medications – Participants were asked to report any changes in their pain medication during the study.

2.5.6 Quality of life – Participants completed the European Quality of Life 5-dimension, 5-level scale (EuroQoL-5D-5L) [29; 33]. Five dimensions are assessed including mobility, self-care, usual activities, pain/discomfort, and anxiety and depression. Each item is scored from 1-5 (1=no problems; 5=unable to/extreme problems). Participants also reported their overall health condition using a visual analogue scale from 0-100 (0 = the worst health you can imagine and 100 = the best health you can imagine. There are no country-specific index values published for Finnish populations.

2.5.7 Adverse events – we collected treatment-related adverse events, serious treatment-related adverse events, adverse device effects, and serious adverse device effects.

- Adverse events were defined as any untoward medical occurrence, unintended disease or injury, or any untoward clinical signs, whether or not related to the investigational medical device. Adverse events were categorized as mild (discomfort noticed, but does not affect normal

activity), moderate (discomfort sufficient to reduce or affect normal daily activity), or severe (incapacitating with inability to work or perform normal daily activities)

- Serious adverse events were defined as any event leading to death or serious deterioration in health of the participant, or fetal distress, death, or congenital abnormality or birth defect.

- Adverse device effects were defined as any adverse effect related to the use of an investigational medical device.

- Serious adverse device effects were defined as any effect that has resulted in any of the consequences characteristic of an SAE.

2.5.8 Patient Clinical Global Impression of Change – Participants were asked to report their global impression of change on a scale from 1-7 (1 = very much improved, 7 = very much worse).

2.5.9 Feedback on the software – The Game Experience Questionnaire [32; 62] was completed by participants after the first VR exposure and at the end of intervention. Participants completed the original 14 items, together with the modified measure, which is specific to low back pain. These questionnaires have previously been used to evaluate VR treatments in healthcare settings (e.g., [3; 8])

2.7 Data analysis

Statistical analyses were conducted in SAS (version 9.4) software. The primary end point was post-treatment, 6 to 8 weeks after randomization. The power calculation for the study considered the standard deviation for the TSK, which is 8 for participants with lower back pain [52]. Thus, with n=10 the precision reached in terms of 95% confidence interval is approximately ±5 units for each group's mean values. This was

considered sufficient precision for getting a preliminary information of the effects on kinesiophobia.

We conducted intention-to-treat (ITT) analyses. All randomized participants with completed baseline questionnaires were entered into ITT analyses. We used a mixed model for repeated measures (MMRM) which can provide unbiased estimates from incomplete data utilizing subject level random effects. When analyzing data, we found 15% missing at 6-8 weeks across outcomes.

We used descriptive statistics to summarize demographics of the sample. Where categorical data are available, we report frequency statistics and where continuous data were collected, we report means and standard deviations. We did not specify a primary hypothesis for the study. Descriptive data are reported for all questionnaires.

As outlined earlier, at the end of the first treatment phase, participants in the standard care group for the first treatment phase were re-randomized to either the DTxP or sham placebo group. We analyzed all participants who had completed DTxP arm, regardless of time of entry into the arm. This means that participants who entered the treatment arms after a period of standard care are combined with participants who had received the treatment in the first phase. Participants who had participated in the standard care intervention first were asked to complete new baseline measures before starting DTxP or sham placebo control treatment in the second phase, and we used those in our analyses.

The TSK and pain intensity NRS outcomes were analyzed using a mixed model for repeated measures (MMRM) to assess change from baseline. Models included fixed effects for planned treatment, randomization strata, study visit and a visit by treatment interaction, random effects for the participant, and baseline scores for each

outcome were included as a covariate. Efficacy was evaluated based on the fitted model, using contrasts to obtain an estimate of difference between treatments.

The ODI and PROMIS measures outcomes were analyzed using an ANCOVA. The models included fixed effects for planned treatment and randomization strata; and the baseline scores as a covariate. Patient clinical global impression of change was analyzed using an ANOVA including fixed effects for planned treatment and randomization strata. A post-hoc Cochran-Mantel-Haenszel analyzing PGIC is reported in Appendix E. Changes to the number of medications are described. No between group analysis was planned for quality of life, and therefore, we only describe this outcome. Changes from baseline to post-treatment are reported for all outcomes.

We planned several subgroup analyses based on an expansion protocol that allowed for up to 100 participants to be recruited into the study. Subgroup analyses were only planned for the TSK outcome, based on age (<55 years vs. $\geq$55 years of age), sex, TSK scores (<42 vs. $\geq$42), NRS pain intensity (<6 vs. $\geq$6), ODI (35 vs. $\geq$35), Keele StarT Back (high, medium, low), CLBP ongoing (<1, 1 vs 5, >5) and are reported in Appendix F. Additional participants were not recruited, so subgroup analyses should be interpreted cautiously.

3.0 Results

3.1 Participants

We assessed 72 participants for eligibility. Thirty-four met the inclusion criteria and we randomized 11 participants to the DTxP, 12 to placebo sham, and 11 to standard care (Figure 4). In the second randomization and allocation, 8 participants who were allocated to standard care were re-randomized to either the DTxP or the sham placebo group. For the blinded comparison between DTxP and the sham placebo there are 42 participants included in the ITT analyses; five were male and 37

were female (mean age 54.69, SD 10.20). All participants were white, and one participant identified as Hispanic or Latino. Participants reported an average pain intensity (NRS) of 5.56 (SD 1.75) at baseline. Two participants had low back pain between 6-12 months, seven reported back pain between 1-5 years, and 33 participants reported back pain lasting for longer than 5 years. See Table 1 for participant demographics.

3.2 Outcome variables

Table 2 shows means and standard deviations of measures at the relevant time-points and between group differences are reported in Table 3. 9-week follow-up data are reported in Appendix G.

3.2.1 Disability: A repeated measures ANCOVA did not show an interaction between treatment and visit (p=0.42) and no simple main effects differences were found between groups at any time-point for ODI scores.

Participants in the DTxP group reported lower ODI scores post-treatment (p=0.05) and end of study (p=0.04) compared to baseline. Participants in the sham placebo comparator did not reduce their ODI scores at post-treatment (p=0.53) but did report lower ODI scores at the end of study compared to baseline (p<0.01). Standard care did not result in lower ODI scores compared to baseline at post-treatment (p=0.66).

A post-hoc analysis of responders to treatment using the ODI cut-offs (responders categorised a-priori as reporting ≤22) found 6/12 participants in the DTxP arm, 1/13 in the sham placebo comparator, and 1/10 in the standard care group were categorised as responders on the ODI at post-treatment. At follow-up, 5/11 participants in the DTxP arm and 2/12 participants in the placebo control were

categorised as responders on the ODI. Due to the inclusion criteria, no participants scored <21% at baseline.

3.2.2 *Pain interference:* When participants in the DTxP group were compared with the sham placebo comparator post-treatment, participants receiving DTxP reported lower pain interference (p=0.05). No differences were found between DTxP and standard care, or the sham placebo comparator group and standard care post-treatment.

Participants in the DTxP group reported lower pain interference (PROMIS 6b) post-treatment compared to baseline (p<0.01). Participants in the sham placebo comparator (p=0.39) and standard care intervention (p=0.22) did not report lower scores compared to baseline. This outcome was not assessed at 5-months follow-up.

3.2.3 *Fear of movement or re-injury:* A repeated measures ANCOVA did not reveal an interaction of treatment by visit (p=0.05). However, simple effects were revealed. When compared to the sham placebo and standard care groups post-treatment, participants in the DTxP arm reported lower TSK scores. There was no difference between the sham placebo comparator and standard care post-treatment. When comparing DTxP and the sham placebo comparator at 5-month follow-up, there were no differences between groups (Table 3; Figure 5).

Participants in the DTxP group reported lower TSK scores at post-intervention (6-8 weeks; p<0.001) and 5-month follow-up (p<0.01) compared to baseline. The sham placebo control did not reduce TSK scores at post-treatment (p=0.29) but did at follow-up (p=0.02) when compared to baseline. The standard care did not reduce TSK scores at post-intervention (p=0.84).

In post-hoc analysis we determined responders to treatment as those participants who reduced their TSK scores by 4 points, which may be considered a

clinically relevant change.[68] We found 10/12 (83%) participants in the DTxP arm, 5/13 (38%) in the sham placebo comparator, and 2/10 (20%) in the standard care group were responders post-treatment. At 5-month follow-up, 8/11 (73%) in the DTxP arm and 7/12 (58%) in the sham placebo comparator arm were categorised as responders.

3.2.4 NRS pain intensity: We did not find an interaction effect for treatment vs. visit (p=0.84). When comparing groups, we found no simple main effects for differences between groups at any time point for average pain intensity.

Participants in the DTxP group reported lower average pain intensity post-treatment compared to baseline (p=0.02), but not at the end of study (p=0.06). Participants in the sham placebo comparator and standard care arms did not report lower average pain intensity post-treatment (p=0.18; p=0.12, respectively), but the sham placebo comparator resulted in lower average pain intensity at 5 month follow-up compared to baseline (p=0.02).

For maximum and minimum pain, there were no interaction effects (p=0.90; p=0.85 respectively). Similarly, there were no simple main effects for any between group comparisons at any time-point when compared to baseline scores for minimum or maximum pain scores.

3.2.5 PROMIS 3a pain intensity: There were no between-group differences between the three groups post-treatment.

Participants in the DTxP group reported lower pain intensity at post-treatment compared to baseline (p<0.01), while participants in the sham placebo comparator (p=0.06) and standard care groups (p=0.07) did not. This measure was not assessed at 5-month follow-up.

3.2.6 Changes in medication: Similar numbers of participants across the three groups were taking medications across the different classes of medications at baseline. Across all groups, 62% of participants were receiving NSAIDS, 50% were receiving opioids, 64% were taking paracetamol, 40% gabapentin or pregabalin, and 17% antidepressants. Two participants in the DTxP group, five participants in the sham placebo comparator, and one participant in the standard care group increased their opioids during the study. For participants taking NSAIDs, three participants in the DTxP group, four in the sham placebo comparator, and two in the standard care group increased their medications during treatment. Three participants in the DTxP group and two in the sham placebo comparator increased their paracetamol consumption, but none in the standard care group reported an increase. Participants did not increase any other medications during treatment.

3.2.7 Quality of life: Quality of life post-treatment scores were not analysed between groups as determined in our protocol. No significant changes were noted for any group from baseline to post-treatment. The DTxP group scored lowest on the mobility (lower scores indicate better quality of life for that domain), pain/discomfort, and self-care domain, whilst the standard care group reported lowest scores on usual activities and anxiety/depression post-treatment. Results were comparable across the three groups on the five domains. The DTxP group reported the highest health state across all three groups post-treatment (higher = better health on this domain).

3.2.8 Adverse events: Thirty-three participants reported 140 adverse events during the treatment period. Eleven participants reported 58 AEs considered relevant to treatment; participants in the sham placebo comparator reported the most treatment-related AEs, followed by the DTxP group, and then the standard care group (Table 4). A small number of participants reported AEs classed as severe (but not

serious adverse events) in each group (DTxP 6 participants reported 8 events; sham placebo comparator 5 participants reported 7 events; standard care 4 participants reported 7 events). These included musculoskeletal and connective tissue disorders (e.g., arthralgia, back pain, and pain in extremity), nervous system disorders (e.g., headache), and motion sickness. There were no serious adverse events (SAE) during the treatment phase.

During follow-up, there was one SAE (back pain) which occurred in the sham placebo group. This was considered unrelated to treatment and the subject recovered. At follow-up, 19 participants reported 44 AEs; 9 participants reported 12 mild AEs, 11 participants reported 18 moderate AEs and 10 participants reported 14 severe AEs (Table 4).

Seven participants reported 20 device deficiencies in the DTxP group, and 6 participants reported 8 device deficiency events in the sham placebo comparator group. These included problems such as a connection problem, bright flashes when using the software, and the program stopping unexpectedly. One participant reported hardware issues of stripes on the eyeglasses in the sham placebo comparator group and two participants in the DTxP group reported either that the hand controllers did not work or there was no voice when the program played.

3.2.9 Patient global impression of change: Participants in the DTxP group reported improved PGIC compared to the sham placebo comparator (p=0.04) and standard care post-treatment (p=0.04), but not compared to the sham placebo comparator group at the end of study (p=0.98). The sham placebo control and standard care groups did not differ post-treatment (p=0.81).

3.2.10 Game experience questionnaire and Modified game experience questionnaire: Participants in the DTxP group rated the VR treatment more positively

compared to the sham placebo comparator group post-treatment (Figure 6 and 7). Participants in the sham placebo comparator group endorsed negative items (e.g., I felt bored, I felt frustrated) more compared to the DTxP group.

3. Discussion

DTxP was superior to both a sham placebo comparator and standard care control in reducing fear of movement and re-injury and improving the global impression of change post treatment. There were no between group differences post-treatment or at follow-up for the remaining outcomes. The within-subject results from this double blind randomized controlled trial show that for adults with moderate-to-severe chronic pain, who are highly disabled and have a high fear of movement and re-injury, that a virtual reality psychological intervention (DTxP) reduced fear of movement, self-reported disability, pain intensity, and pain interference from baseline measures. Participants were generally positive about the VR experience, and no exceptional treatment related adverse events were reported. These results add to the emerging RCT evidence base for the potential utility of VR as a tool to deliver psychological interventions in a physical rehabilitation context [4; 12; 23-25; 61]. This is the first such demonstration of a fully automated system.

Interpretation of these findings should be made within context. First, this is a clinically relevant population with longstanding chronic pain (75% with pain lasting longer than 5 years), who were disabled, had a high fear of movement, and had an average pain intensity of 5.56 out of 10 (SD 1.75) at baseline. Change in such a chronic population is rare. Second, this is a placebo-controlled trial of a psychological intervention with a sham intervention comparator. Controls over blinding were extensive and significant effort was made to ensure that patients remained blind to allocation, and study investigators were blind throughout. Finally, with so few studies

available in the field, we designed the study as a pilot by including multiple measures of interest. However, the active content of the experimental treatment (DTxP) was designed to simultaneously alter beliefs about movement and increase movement – exposing participants to disconfirmatory evidence of repeated goal-relevant movement without pain or injury. Compared to no treatment or sham treatment this manipulation was successful. The study also provides a wealth of interesting data (see appendices) to support further development and testing, including information on treatment content, subgroup analyses, and changes in judgements of global impression of change.

Taking these features into account, this study demonstrates the potential benefit and safety of VR delivered interventions, adding further empirical support to much theorized possible uses in chronic pain functional rehabilitation. There is a mature field of VR based interventions for distraction in procedural pain [34] but chronic pain studies are at an early stage of development. In November 2021, the first FDA approval was given for any VR intervention in chronic pain, for EaseVRx [21,22], recognizing its effects in reducing pain, and in doing so the FDA also created a new regulatory device classification [18] which may significantly reduce barriers to regulatory approval of VR interventions.

To develop the empirical base of VR based interventions in chronic pain, and to increase their potential efficacy in reducing pain, disability and distress, further research is needed in four main areas: 1) optimizing behavioral instruction for use in VR, 2) optimizing exposure and graded activity content for use in VR-rehabilitation, 3) further automation, and 4) innovation in empirical evaluation.

First, early discussion of the use of technology to deliver behavior change instruction was largely focused on the correspondence between machine delivered

and human delivered, holding human delivered intervention as a gold standard. Arguably, current developments in now pervasive high powered portable computing and remote internet access provide tools that can fundamentally change treatments. The emerging field of digital therapeutics encourages a focus on the novel aspects of communication and sensing technologies, in which computing and behavioral content are merged rather than one being the carrier for the other.[40; 44] In VR this means that primary experimental research is critically needed into the behavior change properties, or the mechanism of action, of VR in altering experience. There are excellent examples already on immersion, illusion, and affect regulation, but more are needed. In particular, the extent to which one can provide guided exploration through use of a coach or mentor is unknown. The existing studies are focused in different clinical areas with different content, so direct comparison is difficult. Needed will be guidance on how to develop content, on necessary vs desirable content, and on a usable taxonomy of content to facilitate precision and comparison. The framework provided by Trost and colleagues provides a helpful starting point [63].

Second, interventions that promote movement through exposure need further experimental investigation. Unknown is the optimal grading, order, type, intensity, and presentation of movement tasks applicable in VR. Although people with chronic pain report that after education they valued learning that '*pain does not mean my body is damaged*' [37] by definition fear maintains avoidance and voluntary exposure to feared movement is rare. There is some guidance, however, not least in trials of therapist delivered exposure and graded activity in chronic pain, that can help.[27; 50; 65] Similarly, there is discussion, guidance, and other investigators exploring possible uses in uncontrolled studies or in protocol.[21; 22; 28] Further questions relate to the extent of gamification needed to ensure compliance with repetitive tasks.[57]

Speculatively, the role of attentional engagement may also be important in chronic as well as acute pain environments, to the extent that momentary engagement in goal-orientated tasks may provide temporary relief from pain and facilitate cognitive enrichment.

Third, an important feature of the solution explored here was its automaticity. Removing an interpreting therapist is not a new idea. There are many self-help manuals, audio and video recordings, 'apps', and telephone-based systems available for chronic pain patients. There are also research attempts in cognate fields to more directly mirror successful therapeutic behavior in a non-human agent, both disembodied (e.g., [10]), and embodied; see for example social robots in which a physical agent operates in the real world providing advice, education or comfort (e.g., [51]. To create an embodied non-human agent who can guide behavior change needs empirical guidance. The verbal dynamics of content has always been important (e.g., [16]), but in this environment controlling the nonverbal will be as important, including vocal dynamics (order, tone, pause, affect),[45] facial expression,[47] and posture.[26] Further, in VR the design of both the visual environment and the soundscape to support behavior change content become crucial.[9; 58] Interesting for future studies will be to extend our methods to include a measure of engagement with automatic agents, exploring their credibility, believability, and participant willingness to be guided automatically.

Finally, further study will need to address core challenges in VR research. VR is a complex intervention, and research designs will need to reflect the multiple stakeholders in the process and the inherent variability in practice and content.[55] Relatedly, because computing develops faster than the traditional research cycle much of what is reported in the medical and scientific press, including this report, will

likely discuss technology that will have been superseded, at the time this is being read. Needed is a forum for fast reporting and updating; a just-in-time collaborative linked laboratory model. This is developing in paediatric pain and VR.[38] Complexity and technological transience do not remove the need for high quality and unbiased evaluation – arguably they amplify that need – but the randomized controlled trial alone is a poor fit for establishing efficacy and safety in digital therapeutics. Helpful would be the addition of idiographic research using single case experiment, linguistic analysis, and bioengineering, and further nomothetic research using large-scale real-time monitoring of mass use to allow for machine learning to establish which content supports behavior change and how safety is maintained.

Interpretation of these findings are of course limited by our design features: designed as a pilot the sample is small meaning that type II errors are likely [43]; although we powered this study on changes in fear of movement and re-injury, as a pilot we deliberately did not identify a primary outcome, interested in the direction of multiple endpoints; and we have data only to the end of the study. We did not formally assess patient beliefs about allocation which could be investigated in future studies. Further, we only asked participants to report increases in medication use and so have no data on treatment behavior outside of the interventions. Because of the Covid-19 pandemic, unplanned changes in protocol delivery were required, although randomization and allocation blinding were maintained throughout.

In conclusion, the results of this clinical trial add to the emerging field of virtual reality interventions for chronic pain management and rehabilitation. Replication is necessary. Future invention could usefully develop embodied non-human agents to guide behavior change, make more use of the immersive properties of the environment, and further automate that content to allow for scalable use. Innovation

is also needed in research methods for evaluating the efficacy and safety of complex digital therapeutics.

References

[1] Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, Cella D, Rothrock N, Keefe F, Callahan L, Lai JS. Development of a PROMIS item bank to measure pain interference. Pain 2010;150(1):173-182.

[2] Baker DP, P. Fairbank, J. The Oswestry Disability Index revisited. In: M Roland, J Jenner, editors. Back pain: New Approaches to Rehabilitation and Education Manchester University Press: Manchester, 1989. pp. 174-186.

[3] Bank PJM, Cidota MA, Ouwehand PEW, Lukosch SG. Patient-Tailored Augmented Reality Games for Assessing Upper Extremity Motor Impairments in Parkinson's Disease and Stroke. J Med Syst 2018;42(12):246.

[4] Birckhead B, Eberlein S, Alvarez G, Gale R, Dupuy T, Makaroff K, Fuller G, Liu X, Yu KS, Black JT, Ishimori M, Venuturupalli S, Tu J, Norris T, Tighiouart M, Ross L, McKelvey K, Vrahas M, Danovitch I, Spiegel B. Home-based virtual reality for chronic pain: protocol for an NIH-supported randomised-controlled trial. BMJ Open 2021;11(6):e050545.

[5] Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, Bullock K, Spiegel B. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. JMIR Ment Health 2019;6(1):e11973.

[6] Borsook D, Youssef AM, Simons L, Elman I, Eccleston C. When pain gets stuck: the evolution of pain chronification and treatment resistance. Pain 2018;159(12):2421-2436.

[7] Chen WH, Revicki DA, Amtmann D, Jensen MP, Keefe FJ, Cella D. Development and analysis of PROMIS pain intensity scale, Proceedings of the 18th Annual Conference of the International Society for Quality of Life Research, Vol. 20: Quality of Life Research, 2012. p. 18.

[8] Cidota MAB, P. J. M. Ouwehand, P. W. Lukosch, S. G. Assessing Upper Extremity Motor Dysfunction Using an Augmented Reality Game . . *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017. pp. 144-155.

[9] Cochrane K, Loke L, Leete M, Campbell A, Ahmadpour N. Understanding the First Person Experience of Walking Mindfulness Meditation Facilitated by EEG Modulated Interactive Soundscape, Proceedings of the *Association for Computing Machinery Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction* (*TEI '21*), Vol. 18, 2021. pp. 1-17.

[10] Cristea IA, Sucala M, David D. Can you tell the difference? Comparing face-toface versus computer-based interventions. The "Eliza" effect in psychotherapy. Journal of Cognitive and Behavioral Psychotherapies 2013;13:291-299.

[11] Crombez G, Eccleston C, Van Damme S, Vlaeyen JWS, Karoly P. Fear-avoidance model of chronic pain: the next generation. The Clinical Journal of Pain 2012;28(6):475-483.

[12] Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-Administered Skills-Based Virtual Reality Intervention for Chronic Pain: Randomized Controlled Pilot Study. JMIR Form Res 2020;4(7):e17293.

[13] Eccleston C. Embodied: The Psychology of Physical Sensation. Oxford: Oxford University Press, 2016.

[14] Eccleston C. Chronic pain as embodied defence: implications for current and future psychological treatments. Pain 2018;159 Suppl 1:S17-S23.

[15] Eccleston C, Blyth FM, Dear BF, Fisher E, Keefe FJ, Lynch ME, Palermo TM, Carrington Reid M, de C Williams AC. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. PAIN 2020;*161*(5):889.

[16] Ewbank MP, Cummins R, Tablan V, Bateup S, Catarino A, Martin AJ, Blackwell AD. Quantifying the Association Between Psychotherapy Content and Clinical Outcomes Using Deep Learning. JAMA Psychiatry 2020;77(1):35-43.

[17] Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000;25(22):2940-2952; discussion 2952.

[18] FDA. FDA Authorizes Marketing of Virtual Reality System for Chronic Pain Reduction, 16/11/2021.

[19] Fisher E, Law E, Dudeney J, Palermo TM, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database of Systematic Reviews 2019(4).

[20] Fisher E, Villanueva G, Henschke N, Nevitt SJ, Zempsky W, Probyn K, Buckley B, Cooper TE, Sethna N, Eccleston C. Efficacy and safety of pharmacological, physical, and psychological interventions for the management of chronic pain in children: a WHO systematic review and meta-analysis. Pain 2021.

[21] Fowler CA, Ballistrea LM, Mazzone KE, Martin AM, Kaplan H, Kip KE, Murphy JL, Winkler SL. A virtual reality intervention for fear of movement for Veterans with chronic pain: protocol for a feasibility study. Pilot Feasibility Stud 2019;5:146.

[22] Fowler CA, Ballistrea LM, Mazzone KE, Martin AM, Kaplan H, Kip KE, Ralston K, Murphy JL, Winkler SL. Virtual Reality as a Therapy Adjunct for Fear of Movement in Veterans With Chronic Pain: Single-Arm Feasibility Study. JMIR Form Res 2019;3(4):e11266.

[23] France CR, Thomas JS. Virtual immersive gaming to optimize recovery (VIGOR) in low back pain: A phase II randomized controlled trial. Contemp Clin Trials 2018;69:83-91.

[24] Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, Salmasi V, Maddox T, Darnall BD. An 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality Program for Chronic Low Back Pain: Double-Blind, Randomized, Placebo-Controlled Trial Conducted During COVID-19. J Med Internet Res 2021;23(2):e26292.

[25] Garcia LM, Darnall BD, Krishnamurthy P, Mackey IG, Sackman J, Louis RG, Maddox T, Birckhead BJ. Self-Administered Behavioral Skills-Based At-Home Virtual Reality Therapy for Chronic Low Back Pain: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2021;10(1):e25291.

[26] García E, Di Paolo EA, De Jaegher H. Embodiment in online psychotherapy: A qualitative study. Psychol Psychother 2021.

[27] Glombiewski JA, Holzapfel S, Riecke J, Vlaeyen JWS, de Jong J, Lemmer G, Rief W. Exposure and CBT for chronic back pain: An RCT on differential efficacy and optimal length of treatment. J Consult Clin Psychol 2018;86(6):533-545.

[28] Griffin A, Wilson L, Feinstein AB, Bortz A, Heirich MS, Gilkerson R, Wagner JF, Menendez M, Caruso TJ, Rodriguez S, Naidu S, Golianu B, Simons LE. Virtual Reality in Pain Rehabilitation for Youth With Chronic Pain: Pilot Feasibility Study. JMIR Rehabil Assist Technol 2020;7(2):e22620.

[29] Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20(10):1727-1736.

[30] Hoffman HG. Interacting with virtual objects via embodied avatar hands reduces pain intensity and diverts attention. Sci Rep 2021;11(1):10672.

[31] Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE, Dixon-Woods M, McCulloch P, Wyatt JC, Chan AW, Michie S. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687.

[32] IJsselsteijn WA, de Kort YAW, Poels K. T*he Game Experience Questionnaire*. Technische Universiteit Eindhoven, 2013.

[33] Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013;22(7):1717-1727.

[34] Keefe FJ, Huling DA, Coggins MJ, Keefe DF, Rosenthal ZM, Herr NR, Hoffman HG. Virtual reality for persistent pain: a new direction for behavioral pain management. Pain 2012;153(11):2163-2166.

[35] Koho P, Aho S, Kautiainen H, Pohjolainen T, Hurri H. Test-retest reliability and comparability of paper and computer questionnaires for the Finnish version of the Tampa Scale of Kinesiophobia. Physiotherapy 2014;100(4):356-362.

[36] Ku B, Lupton E. Health Design Thinking: Creating Products and Services for Better Health. New York: Cooper Hewitt, Smithsonian Design Museum, 2020.

[37] Leake HB, Moseley GL, Stanton TR, O'Hagan ET, Heathcote LC. What do patients value learning about pain? A mixed-methods survey on the relevance of target concepts after pain science education. Pain 2021;162(10):2558-2568.

[38] Logan DE, Simons LE, Caruso TJ, Gold JI, Greenleaf W, Griffin A, King CD, Menendez M, Olbrecht VA, Rodriguez S, Silvia M, Stinson JN, Wang E, Williams SE, Wilson L. Leveraging Virtual Reality and Augmented Reality to Combat Chronic Pain in Youth: Position Paper From the Interdisciplinary Network on Virtual and Augmented Technologies for Pain Management. J Med Internet Res 2021;23(4):e25916.

[39] Ma L, Mor S, Anderson PL, Baños RM, Botella C, Bouchard S, Cárdenas-López G, Donker T, Fernández-Álvarez J, Lindner P, Mühlberger A, Powers MB, Quero S, Rothbaum B, Wiederhold BK, Carlbring P. Integrating virtual realities and psychotherapy: SWOT analysis on VR and MR based treatments of anxiety and stress-related disorders. Cogn Behav Ther 2021;50(6):509-526.

[40] Makin S. The emerging world of digital therapeutics. Nature 2019;573(7775):S106-S109.

[41] Miller RP, Kori SH, Todd DD. The Tampa Scale: a Measure of Kinisophobia. The Clinical Journal of Pain 1991;7(1):51.

[42] NICE. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. NG193., 2021.

[43] Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Juni P. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. Bmj 2010;341:c3515.

[44] Patel NA, Butte AJ. Characteristics and challenges of the clinical pipeline of digital therapeutics. NPJ Digit Med 2020;3(1):159.

[45] Paz A, Rafaeli E, Bar-Kalifa E, Gilboa-Schectman E, Gannot S, Laufer-Goldshtein B, Narayanan S, Keshet J, Atzil-Slonim D. Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. J Consult Clin Psychol 2021;89(3):227-239.

[46] Pekkanen L, Kautiainen H, Ylinen J, Salo P, Hakkinen A. Reliability and validity study of the Finnish version 2.0 of the oswestry disability index. Spine (Phila Pa 1976) 2011;36(4):332-338.

[47] Peluso PR, Freund RR. Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. Psychotherapy (Chic) 2018;55(4):461-472.

[48] Rantonen J, Vehtari A, Karppinen J, Luoto S, Viikari-Juntura E, Hupli M, Malmivaara A, Taimela S. Face-to-face information combined with a booklet versus a booklet alone for treatment of mild low-back pain: a randomized controlled trial. Scand J Work Environ Health 2014;40(2):156-166.

[49] Ridout B, Kelson J, Campbell A, Steinbeck K. Effectiveness of Virtual Reality Interventions for Adolescent Patients in Hospital Settings: Systematic Review. J Med Internet Res 2021;23(6):e24967.

[50] Riecke J, Rief W, Vlaeyen JWS, Glombiewski JA. Generalizability of harm and pain expectations after exposure in chronic low back pain patients. Eur J Pain 2020;24(8):1495-1504.

[51] Robinson NL, Connolly J, Hides L, Kavanagh DJ. Social robots as treatment agents: Pilot randomized controlled trial to deliver a behavior change intervention. Internet Interv 2020;21:100320.

[52] Roelofs J, van Breukelen G, Sluiter J, Frings-Dresen MH, Goossens M, Thibault P, Boersma K, Vlaeyen JW. Norming of the Tampa Scale for Kinesiophobia across pain diagnoses and various countries. Pain 2011;152(5):1090-1095.

[53] Rosser BA, Fisher E, Eccleston C, Duggan GB, Keogh E. Psychological therapies delivered remotely for the management of chronic pain (excluding headache) in adults. Cochrane Database of Systematic Reviews 2021.

[54] Roxburgh T, Li A, Guenancia C, Pernollet P, Bouleti C, Alos B, Gras M, Kerforne T, Frasca D, Le Gal F, Christiaens L, Degand B, Garcia R. Virtual Reality for Sedation During Atrial Fibrillation Ablation in Clinical Practice: Observational Study. J Med Internet Res 2021;23(5):e26349.

[55] Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, Boyd KA, Craig N, French DP, McIntosh E, Petticrew M, Rycroft-Malone J, White M, Moore L. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061.

[56] Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, Albert M, Tashjian V, Lam R, Ahn J, Dailey F, Rosen BT, Vrahas M, Little M, Garlich J, Dzubur E, IsHak W, Danovitch I. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PLoS One 2019;14(8):e0219115.

[57] Steiner B, Elgert L, Saalfeld B, Wolf KH. Gamification in Rehabilitation of Patients With Musculoskeletal Diseases of the Shoulder: Scoping Review. JMIR Serious Games 2020;8(3):e19914.

[58] Tabbaa L, Ang CS, Siriaraya P, She WJ, Prigerson HG. A Reflection on Virtual Reality Design for Psychological, Cognitive and Behavioral Interventions: Design Needs, Opportunities and Challenges. International Journal of Human–Computer Interaction 2021;37(9):851-866.

[59] Tabor A, Keogh E, Eccleston C. Embodied pain-negotiating the boundaries of possible action. Pain 2017;158(6):1007-1011.

[60] Tack C. Virtual reality and chronic low back pain. Disabil Rehabil Assist Technol 2021;16(6):637-645.

[61] Tejera DM, Beltran-Alacreu H, Cano-de-la-Cuerda R, Leon Hernández JV, Martín-Pintado-Zugasti A, Calvo-Lobo C, Gil-Martínez A, Fernández-Carnero J. Effects of Virtual Reality versus Exercise on Pain, Functional, Somatosensory and Psychosocial Outcomes in Patients with Non-specific Chronic Neck Pain: A Randomized Clinical Trial. Int J Environ Res Public Health 2020;17(16).

[62] Thomas JS, France CR, Applegate ME, Leitkam ST, Walkowski S. Feasibility and Safety of a Virtual Reality Dodgeball Intervention for Chronic Low Back Pain: A Randomized Clinical Trial. J Pain 2016;17(12):1302-1317.

[63] Trost Z, France C, Anam M, Shum C. Virtual reality approaches to pain: toward a state of the science. Pain 2021;162(2):325-331.

[64] van Hooff ML, Mannion AF, Staub LP, Ostelo RW, Fairbank JC. Determination of the Oswestry Disability Index score equivalent to a "satisfactory symptom state" in patients undergoing surgery for degenerative disorders of the lumbar spine- a Spine Tango registry-based study. Spine J 2016;16(10):1221-1230.

[65] Vlaeyen JW, de Jong J, Geilen M, Heuts PH, van Breukelen G. Graded exposure in vivo in the treatment of pain-related fear: a replicated single-case experimental design in four patients with chronic low back pain. Behav Res Ther 2001;39(2):151-166.

[66] Williams ACC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020;8:CD007407.

[67] Williams ACC, Fisher E, Hearn L, Eccleston C. Evidence-based psychological interventions for adults with chronic pain: precision, control, quality, and equipoise. Pain 2021;162(8):2149-2153.

[68] Woby SR, Roach NK, Urmston M, Watson PJ. Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. Pain 2005;117(1-2):137-144.

Figure Legends

Figure 1. Study design

Figure 2. Exploring the 'cabin' inside space used for the Digital Therapeutics for Pain (DTxP) intervention and the Sham intervention. *

Figure 3. Exploring the outside space used for the Digital Therapeutics for Pain (DTxP) intervention and the Sham intervention. *

*The participants in these images were members of the design team demonstrating the environments, not study participants.

Figure 4. CONSORT Flow diagram.

Figure 5. TSK scores between groups.

Figure 6. Game Experience Questionnaire at post-treatment

Figure 7. Modified Game experience Questionnaire at post-treatment

Figure 1.

KEY:
Treatment A = DTxP
Treatment B = Sham Placebo
R = Randomisation
RR = Re-Randomisation
wks = Weeks
EOI = End-of-Intervention
EOS = End-of-Study

* Study will be unblinded after week 9 follow-up of the last patient, following database lock.
  The open label extended follow-up will continue until 5 months after randomisation (~3 months after EOI)
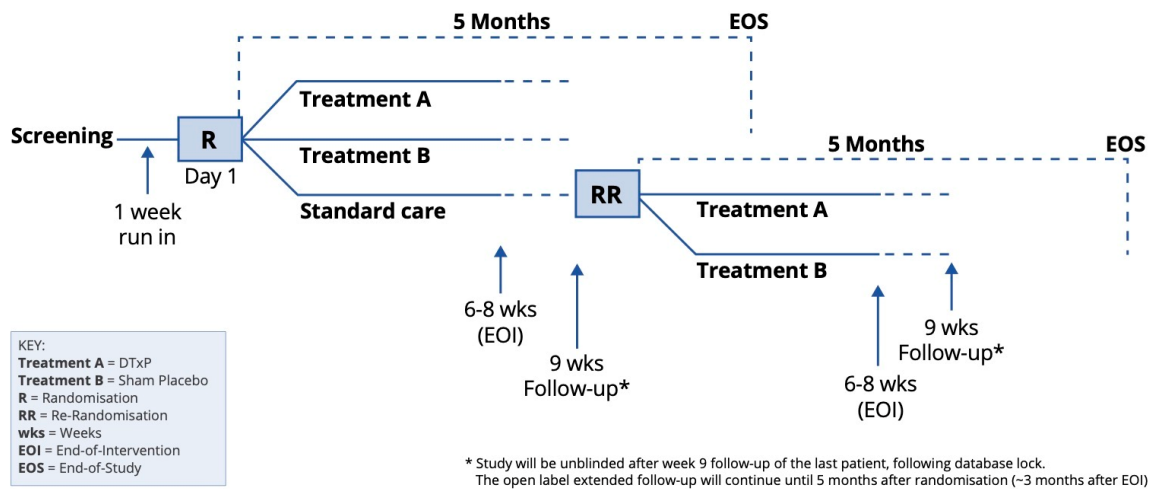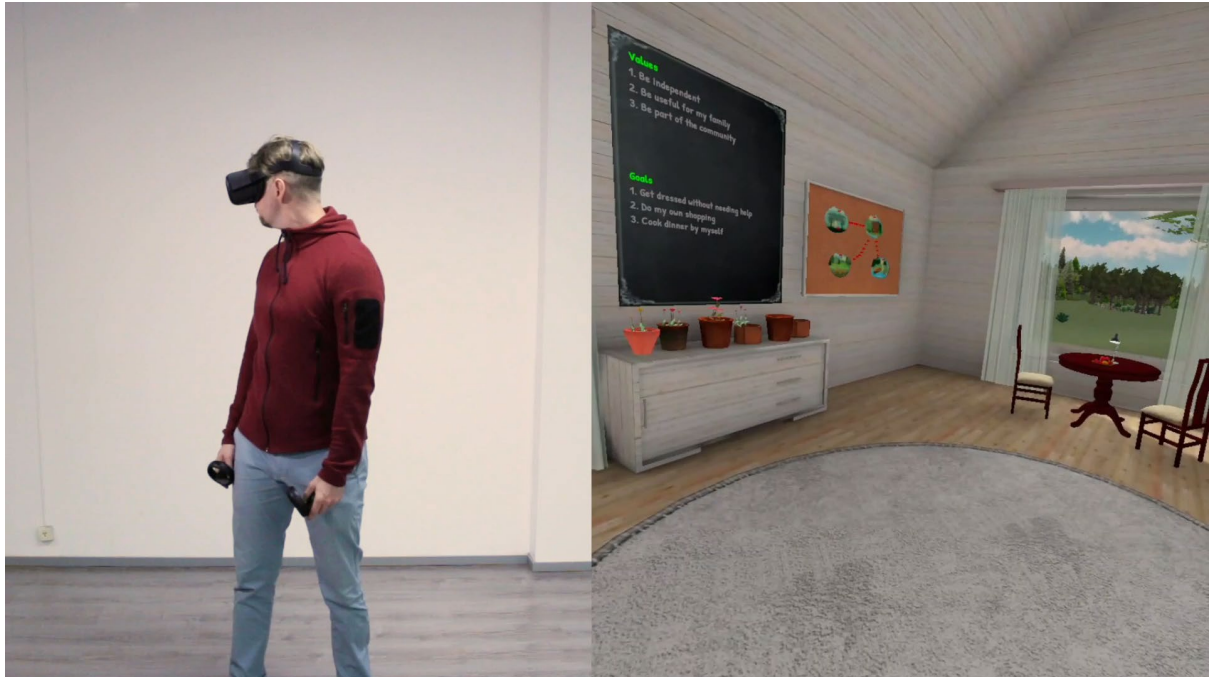
Figure 2

Figure 3

Figure 4
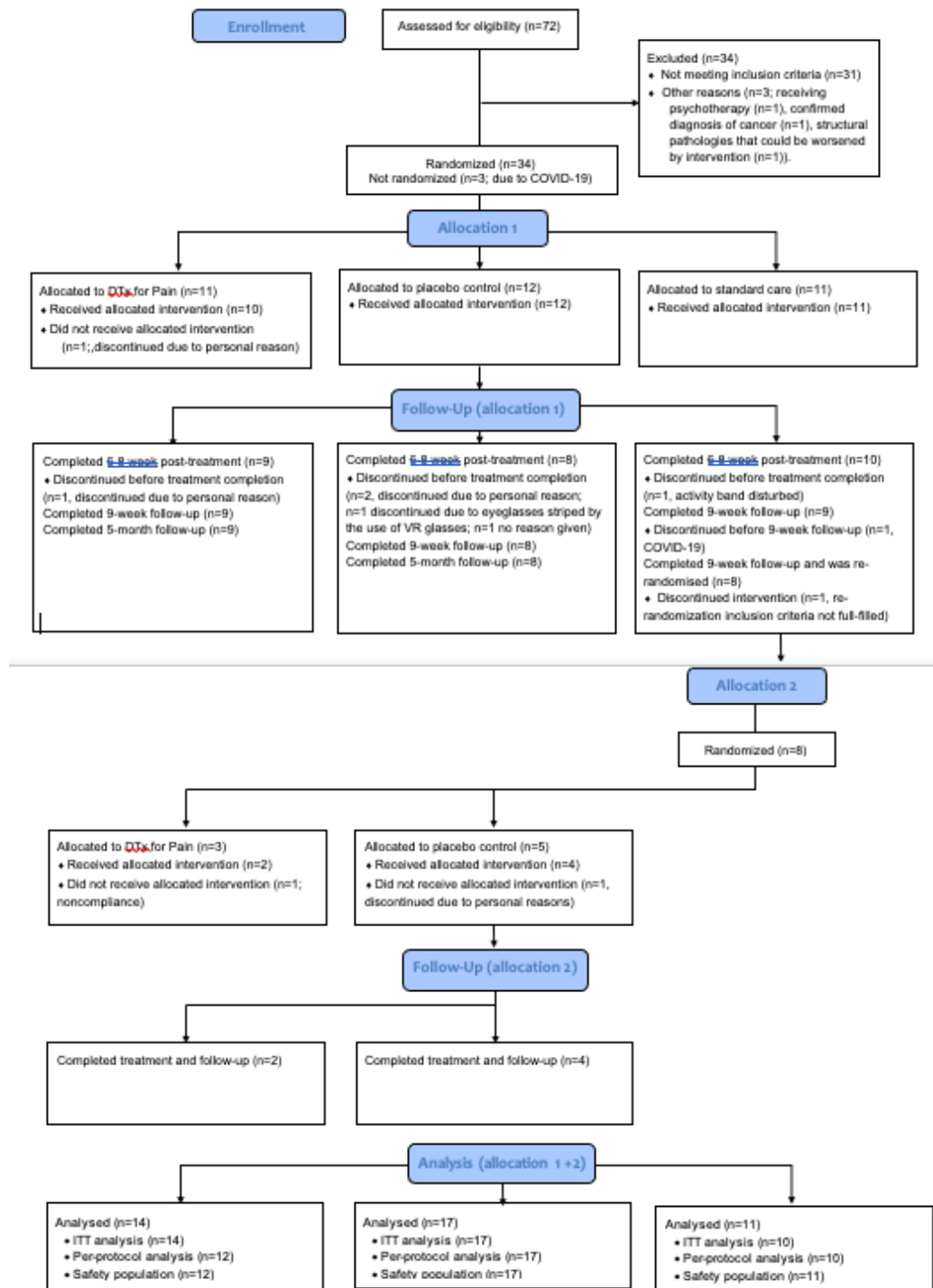
Figure 5
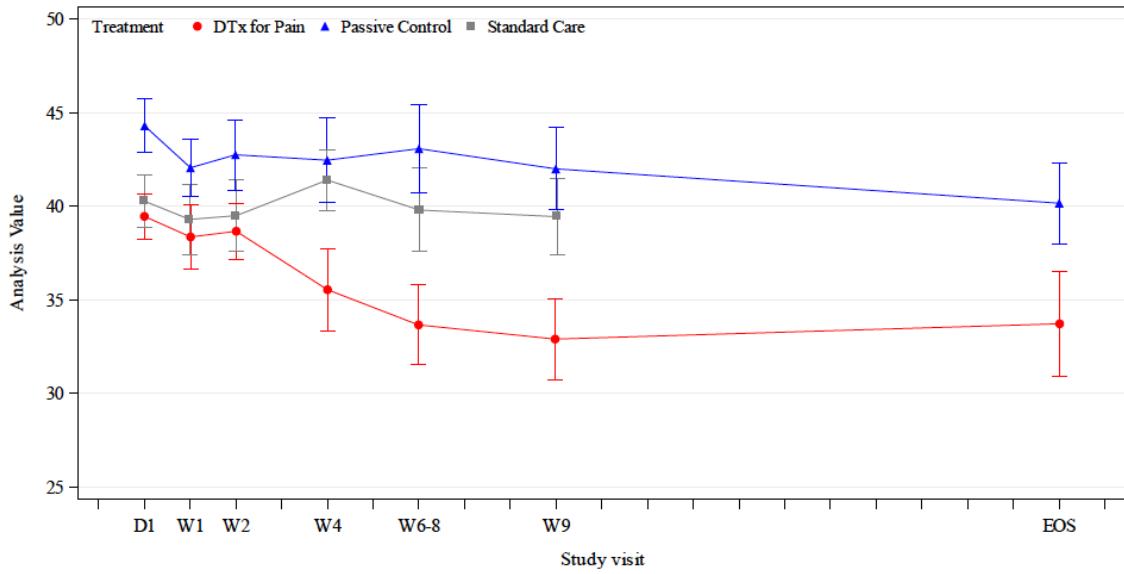
Figure 6

End of Intervention (6-8 Weeks)
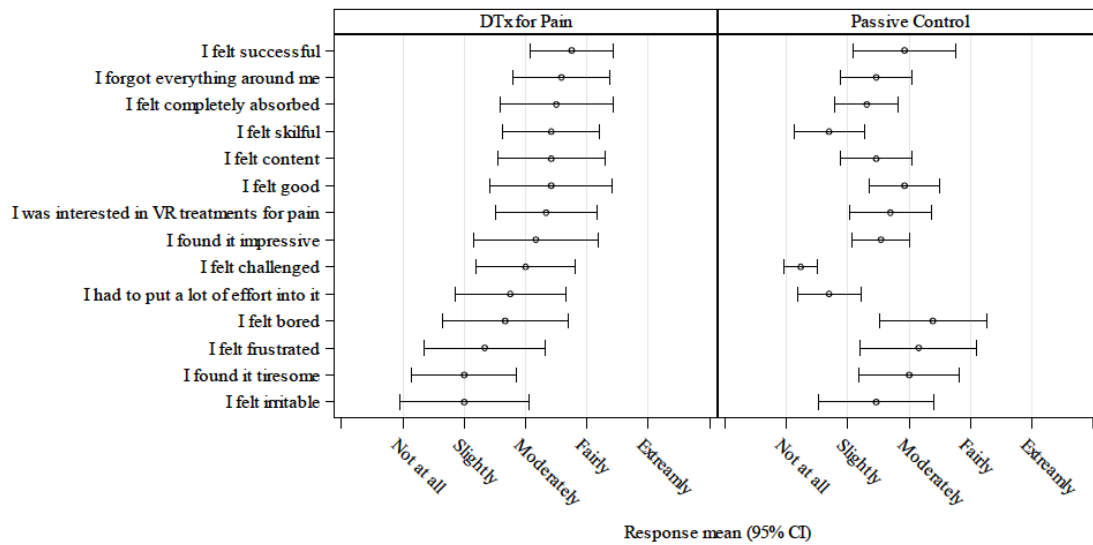
Figure 7

End of Intervention (6-8 Weeks)
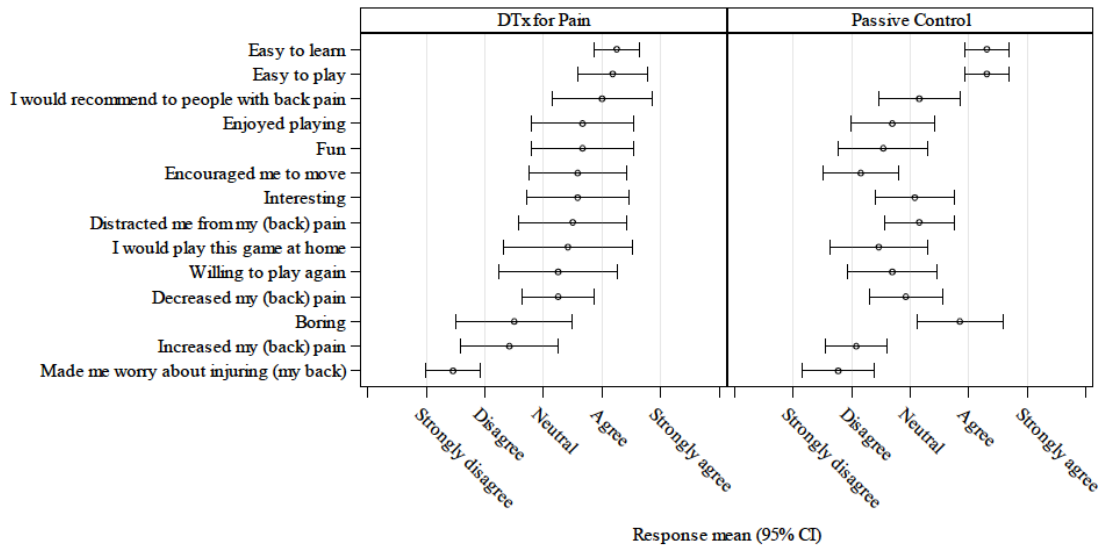
Table 1. Participant demographics at baseline

	DTxP (N=14)	Sham placebo (N=17)	Standard care (N=11)	Total (N=42)
Age (years (SD))	55.14 (10.53)	52.76 (11.19)	57.09 (8.34)	54.69 (10.20)
Sex (M:F)	2:12	3:14	0:11	5:37
Race				
White	14	17	11	42
Ethnicity				
Hispanic or Latino	0	0	1	1
Not Hispanic or Latino	14	17	10	41
Duration of back pain				
6-12 months	2	0	0	2
1-5 years	2	4	1	7
>5 years	10	13	10	33

Table 2. Outcome measures (ITT sample)

	DTxP (N=14; Mean (SD))	Sham placebo (N =17; Mean (SD))	Standard care (N=11; Mean (SD))
Oswestry Disability Index			
Screening	36.0 (7.6)	37.2 (9.4)	36.2 (7.6)
Day 1	34.8 (11.0)	38.2 (9.4)	33.8 (7.5)
Post-treatment	28.8 (15.6)*	38.5 (16.2)	32.8 (8.6)
End of study	29.1 (17.2)*	31.3 (12.8)**	-
Pain interference (PROMIS 6b)			
Screening	64.5 (3.7)	63.1 (3.4)	63.1 (2.5)
Day 1	63.6 (2.3)	63.2 (2.9)	62.1 (3.5)
Post-treatment	59.0 (6.6)**	62.6 (5.3)	60.9 (3.8)
Tampa Scale for Kinesiophobia			
Screening	41.9 (4.4)	43.2 (6.0)	42.5 (5.4)
Day 1	39.5 (4.4)	44.3 (5.9)	40.3 (4.6)
Post-treatment	33.7 (7.4)***	43.1 (8.5)	39.8 (7.1)
End of study	33.7 (9.2)**	40.2 (7.5)*	-
Average pain (NRS)			
Screening	6.0 (1.4)	6.1 (1.4)	5.7 (1.6)
Day 1	5.7 (1.4)	5.6 (1.9)	5.4 (2.0)
Post-treatment	4.1 (1.7)*	4.8 (2.3)	4.4 (2.4)
End of study	4.4 (1.9)	4.0 (1.8)*	-
Maximum pain (NRS)			
Screening	8.2 (0.9)	8.3 (0.9)	8.5 (0.9)
Day 1	7.5 (1.3)	7.5 (1.8)	7.4 (1.6)
Post-treatment	6.0 (2.4)	6.9 (2.5)	6.9 (2.5)
End of study	6.8 (2.1)	6.3 (2.7)*	-
Minimum pain (NRS)			
Screening	3.5 (2.1)	4.2 (2.0)	4.5 (1.9)
Day 1	3.8 (1.7)	4.4 (2.0)	3.6 (2.7)
Post-treatment	2.6 (1.6)	3.5 (2.3)	2.8 (2.5)
End of study	2.7 (1.7)	2.6 (1.7)*	-
Pain intensity (PROMIS 3a)			
Screening	66.5 (4.1)	65.1 (5.4)	63.0 (5.5)
Day 1	66.1 (5.6)	64.5 (5.2)	64.5 (5.9)
Post-treatment	60.0 (7.5)**	61.9 (8.3)	61.0 (5.9)
EQ-5D			
Mobility			
Day 1	2.2 (0.7)	2.5 (0.9)	2.1 (0.5)
Post treatment	2.0 (1.0)	2.3 (0.9)	2.1 (0.6)
Usual activities			
Day 1	2.6 (0.8)	2.2 (0.6)	1.9 (0.5)
Post treatment	2.2 (1.2)	2.2 (0.9)	1.8 (0.6)
Anxiety/Depression			
Day 1	1.7 (0.9)	1.6 (0.7)	1.2 (0.4)
Post treatment	2.0 (1.2)	1.8 (0.8)	1.2 (0.4)
Pain/Discomfort			
Day 1	3.2 (0.4)	2.7 (0.7)	2.8 (0.8)
Post treatment	2.5 (0.8)	2.6 (0.8)	2.6 (0.7)
Self-care			
Day 1	1.5 (0.8)	1.6 (0.6)	1.8 (0.6)
Post treatment	1.3 (0.7)	1.7 (0.6)	1.5 (0.7)
Health state (0-100)			
Day 1	47.7 (16.8)	63.8 (15.1)	55.7 (19.6)

Post treatment	66.1 (22.2)	58.1 (26.8)	62.0 (23.9)
PGIC			
Post-treatment	2.7 (1.4)***	3.8 (1.5)***	3.9 (0.7)***
End of study	3.0 (1.5)***	3.0 (1.5)***	-

M: mean; PGIC: patient global impression of change; SD: Standard deviation; significance levels are compared to day 1, *<0.05, **<0.01, ***<0.001.

Table 3. Treatment effects from repeated measures ANCOVA (ITT sample*)

Outcome/Group/Time-point	Estimate (SE)	95% CIs	P-value
Disability (ODI) post-treatment			
DTxP vs. Sham placebo	-4.86 (4.48)	-13.98 to 4.26	0.29
DTxP vs. Standard care	-5.25 (4.67)	-14.77 to 4.27	0.27
Sham placebo vs. Standard care	-0.39 (4.69)	-9.95 to 9.17	0.93
Disability (ODI) end of study			
DTxP vs. Sham placebo	1.88 (4.16)	-6.76 to 10.51	0.66
Pain interference (Promis 6b) post-treatment			
DTxP vs. Sham placebo	-3.98 (1.96)	-7.99 to 0.03	0.05
DTxP vs. Standard care	-3.20 (2.13)	-7.55 to 1.16	0.14
Sham placebo vs. Standard care	0.79 (2.07)	-3.45 to 5.02	0.71
Fear of movement (TSK) post-treatment			
DTxP vs. Sham placebo	-4.91 (2.30)	-9.57 to -0.25	0.04
DTxP vs. Standard care	-6.16 (2.35)	-10.95 to -1.38	0.01
Sham placebo vs. Standard care	-1.25 (2.36)	-6.04 to 3.53	0.60
Fear of movement (TSK) end of study			
DTxP vs. Sham placebo	-1.74 (2.68)	-7.28 to 3.80	0.52
Average Pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.73 (0.82)	-2.40 to 0.94	0.38
DTxP vs. Standard care	-0.42 (0.90)	-2.25 to 1.41	0.64
Sham placebo vs. Standard care	0.31 (0.86)	-1.44 to 2.06	0.72
Pain intensity (NRS) end of study			
DTxP vs. Sham placebo	0.32 (0.82)	-1.44 to 2.08	0.70
Maximum pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.35 (1.04)	-2.47 to 1.78	0.74
DTxP vs. Standard care	-0.87 (1.14)	-3.20 to 1.47	0.45
Sham placebo vs. Standard care	-0.52 (1.10)	-2.75 to 1.71	0.64
Pain intensity (NRS) end of study			
DTxP vs. Sham placebo	0.89 (0.88)	-0.98 to 2.76	0.33
Minimum pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.88 (0.84)	-2.58 to 0.83	0.30
DTxP vs. Standard care	-0.24 (0.92)	-2.11 to 1.63	0.79
Sham placebo vs. Standard care	-0.63 (0.88)	-1.16 to 2.43	0.48
Pain intensity (NRS) end of study			
DTxP vs. Sham placebo	0.28 (0.81)	-1.46 to 2.03	0.73
Pain intensity (Promis 3a) post-treatment			
DTxP vs. Sham placebo	-2.48 (2.88)	-8.35 to 3.40	0.40
DTxP vs. Standard care	-2.07 (3.09)	-8.39 to 4.24	0.51
Sham placebo vs. Standard care	0.41 (3.01)	-5.75 to 6.56	0.89
PGIC post-treatment			
DTxP vs. Sham placebo	-1.12 (0.54)	-2.22 to -0.03	0.04
DTxP vs. Standard care	-1.26 (0.57)	-2.43 to -0.09	0.04
Sham placebo vs. Standard care	-0.13 (0.56)	-1.28 to 1.01	0.81
PGIC end of study			
DTxP vs. Sham placebo	-0.01 (0.62)	-1.29 to 1.27	0.98

CI: confidence intervals; DTXP: Digital treatment for Pain; NRS: numerical rating scale; ODI: Oswestry Disability Index; PGIC: patient global impression of change; SE: standard error; TSK: Tampa Scale for Kinesiophobia. *DTxP N=14; Sham placebo N=17; Standard care N=11.

Table 4. Summary of adverse events (AE) during treatment (number of participants, % of participants, and event count)

Safety population

Adverse event type	DTxP (N=12) Participant N, (%), Event N	Sham placebo (N=17) Participant N, (%), Event N	Standard Care (N=11) Participant N, (%), Event N	Total (N=40) Participant N, (%), Event N
Post-treatment				
All AEs	12 (100) 50	14 (82.4) 60	7 (63.6) 30	33 (82.5) 140
Mild AEs	4 (33.3) 17	6 (35.3) 17	2 (18.2) 7	12 (30.0) 41
Moderate AEs	10 (83.3) 25	11 (64.7) 36	6 (54.5) 16	27 (67.5) 77
Severe AEs	6 (50.0) 8	5 (29.4) 7	4 (36.4) 7	15 (37.5) 22
Treatment-related AEs	3 (25.0) 25	6 (35.3) 29	2 (18.2) 4	11 (27.5) 58
Serious AEs	0 (0) 0	0 (0) 0	0 (0) 0	0 (0) 0
Follow-up				
All AEs	7 (58.3) 18	9 (52.9) 23	3 (27.3) 3	19 (47.5) 44
Mild AEs	3 (25.0) 5	5 (29.4) 6	1 (9.1) 1	9 (22.5) 12
Moderate AEs	5 (41.7) 6	5 (29.4) 11	1 (9.1) 1	11 (27.5) 18
Severe AEs	4 (33.3) 7	5 (29.4) 6	1 (9.1) 1	10 (25.0) 14
Serious AEs		1 (5.9) 1		1 (2.5) 1

Appendix A: User experience flow over the 30 days, themes and module placement.

Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Theme 1: Orientation	Theme 2: Hurt not Harm	Theme 3: Problem Solving	Theme 4: Plans & Setbacks	Theme 5: Peripersonal space	Theme 6: Consolidation
1.1. Orientation & Introduction	2.1. Brain, Nerves and Mind (M4) (M5)	3.1. Problem solving: cognitive (M8) (M9)	4.1. Barriers to change and Setback planning (M12) (M13)	5.1. Growing in space (M16, M17)	6.1. Escape Velocity (M20, M21)
1.2. Practice	2.2. Practice	3.2. Practice	4.2. Practice	5.2. Practice	6.2. Practice
1.3. Values & Goals (M1) (M2)	2.3. Hurt Not Harm (M6)	3.3. Problem Solving: social (M10)	4.3. Identity (re) narrative (M14)	5.3. Avoidance-Endurance (M18)	6.3. Review of Plans (M22) (M23)
1.4. Practice	2.4 Practice	3.4. Practice	4.4. Practice	5.4. Practice	6.4. Final Game
1.5. Review, Practice and Homework (M3)	2.5. Review, Practice and Homework (M7)	3.5. Review, Practice and Homework (M11)	4.5. Review, Practice and Homework (M15)	5.5. Review, Practice and Homework (M19)	6.5. Time to go forward (M24)
1.W. Homework	2.W. Homework	3.W. Homework	4.W. Homework	5.W. Homework	

Appendix B: User experience flow over the 30 days, themes and module placement.

Module	Name	Module	Name
M1	What matters to me	**M13**	Setback planning
M2	What next	**M14**	Me, myself, I
M3	Homework planning 1	**M15**	Homework planning 4
M4	Homework report 1	**M16**	Homework report 4
M5	Brain nerves and mind	**M17**	Space invaders
M6	Hurt not harm	**M18**	Push me-pull you
M7	Homework planning 2	**M19**	Homework planning 5
M8	Homework report 2	**M20**	Homework report 5
M9	Cognitive problem solving	**M21**	Escape velocity
M10	Social problem solving	**M22**	Virtual futures
M11	Homework planning 3	**M23**	Toolbox
M12	Homework report 3	**M24**	Time to go…forward

Appendix C. Four behavior change themes, description of content, and how the modules map onto them.

Relationship Maintenance	Description	
Rationale	Rationale building, relationship building, compassionate regard	M1, M2
Education	Harm not hurt, pain without injury, behaviour and its consequences	M6
Setback	Guidance on setback planning (overdo, underdo, lose motivation)	M12, M13
Reward	Positive reinforcement for goal & value determined behaviour change?	Throughout
Engage	Material is engaging and makes me want to return	Throughout
Embodied Reactivity	**Description**	
Peripersonal	Frequent behaviour to own and extend peripersonal space	M16,M17
Motion	Range of motion exercises relevant to the indication.	Throughout
Courageous Engagement	**Description**	
Courage	Counter-intuitive experience of feared harms	M4, M5, M18
Emotion	Assessment of emotional reaction in task	M16, M17, M20, M21
Pacing	Contingent reinforcement of goal-determined activities	M1, M2, M22, M23 & Throughout
Mastery	**Description**	
Cognitive	Cognitive Problem Solving (memory, attention, decision-making)	M8, M9
Social	Social Problem Solving (worry checking, problem reframing, confidence)	M10

Appendix D: TiDiER checklist

The TIDieR (Template for Intervention Description and Replication) Checklist*:
Information to include when describing an intervention and the location of the information

Item number	Item	Where located **	
		Primary paper (page or appendix number)	Other † (details)
	BRIEF NAME		
1.	Provide the name or a phrase that describes the intervention.	Page 1	**Title**: A prospective, double-blind, pilot randomized controlled trial of 'embodied' virtual reality intervention for adults with low back pain and high fear of movement.
	WHY		
2.	Describe any rationale, theory, or goal of the elements essential to the intervention.	Page 3	**Introduction:** "The content was maximized to make use of the core properties of virtual reality: immersion, exploration, and experimentation. The intervention requires daily goal-oriented movement that increases in both intensity and range of movement over time [56]. The content is guided by an 'embodied' model of the psychology of pain in which disability and distress arise from a) the privileged access to consciousness of threat-relevant interoception (i.e., bodily sensations are more likely to be attended to, interpreted as threatening, and acted upon), b) avoidance behavior maintained with reinforcement by immediate and general safety signals, and generalized to multiple contexts, and c) consequent social and cognitive failure supported by self-defeating behavior and cognition [11; 12; 55]. Moving without concern for harm, experiencing positive outcomes after self-determined action, and safely experimenting with core beliefs can all be delivered in an immersive environment within VR."
	WHAT		
3.	Materials: Describe any physical or informational materials used in the	Pages 6-10	**Procedures:** "During this session, participants were given a VR wireless headset and hand controllers, corresponding to their

	intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (e.g. online appendix, URL).		randomization number. Participants were shown how to use the headset, hand-held controllers and all movements required by the treatment. The DTxP or sham program was pre-loaded. Early during recruitment, in-person sessions were suspended due to COVID-19 and replaced with training videos, instructions, and phone support. VR headsets were delivered to participants' homes. During home-use, participants were instructed to use the VR technology on a clear, flat ground surface 2m x 2m in size to prevent injury. Participants range of movement and stability were assessed by staff at the first session." More information is provided in the manuscript.
4.	Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities.	Page 6-10, Appendix A-C	Too much information to include in the checklist. Please see treatment description and appendices.
	WHO PROVIDED		
5.	For each category of intervention provider (e.g. psychologist, nursing assistant), describe their expertise, background and any specific training given.	Page 7	This is an automated intervention provided by a digital mentor. The DTxP and Sham placebo intervention were principally designed by Professor Eccleston, who has a UK Health Professionals Council registration to practice health psychology (PYL03729), 15 years' experience in interdisciplinary residential chronic pain management, and has written on the methods of embodied psychology. The protocols, and inclusion and exclusion criteria were reviewed by Professors Rice and Jääskeläinen, and the movements promoted in the activity were reviewed by experts from the Finnish Spinal Health Association
	HOW		
6.	Describe the modes of delivery (e.g. face-to-face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group.	Page 6	Participants were given VR headsets.
	WHERE		

7.	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features.	Page 4	**Procedures:** During home-use, participants were instructed to use the VR technology on a clear, flat ground surface 2m x 2m in size to prevent injury. Participants range of movement and stability were assessed by staff at the first session.
	WHEN and HOW MUCH		
8.	Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity or dose.	Page 7	**Treatment groups:** '*Duration and exposure.* Both the DTxP and sham placebo treatments were designed to last 15-60 minutes per session. There were five sessions scheduled for each week, resulting in 30 unique days that were designed to be delivered over 6-8 weeks.
	TAILORING		
9.	If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how.	N/A	No tailoring was made
	MODIFICATIONS		
10.‡	If the intervention was modified during the course of the study, describe the changes (what, why, when, and how).	N/A	No modifications were made
	HOW WELL		
11.	Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them.	N/A	This is an automated intervention provided by a digital mentor. No alterations were made to the programs.
12.‡	Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned.	N/A	This is an automated intervention provided by a digital mentor.

** **Authors** - use N/A if an item is not applicable for the intervention being described. **Reviewers** – use '?' if information about the element is not reported/not   sufficiently reported.

† If the information is not provided in the primary paper, give details of where this information is available. This may include locations such as a published protocol   or other published papers (provide citation details) or a website (provide the URL).

‡ If completing the TIDieR checklist for a protocol, these items are not relevant to the protocol and cannot be described until the study is complete.

* We strongly recommend using this checklist in conjunction with the TIDieR guide (see *BMJ* 2014;348:g1687) which contains an explanation and elaboration for each item.

* The focus of TIDieR is on reporting details of the intervention elements (and where relevant, comparison elements) of a study. Other elements and methodological features of studies are covered by other reporting statements and checklists and have not been duplicated as part of the TIDieR checklist. When a **randomised trial** is being reported, the TIDieR checklist should be used in conjunction with the CONSORT statement (see www.consort-statement.org) as an extension of **Item 5 of the CONSORT 2010 Statement.** When a **clinical trial protocol** is being reported, the TIDieR checklist should be used in conjunction with the SPIRIT statement as an extension of **Item 11 of the SPIRIT 2013 Statement** (see www.spirit-statement.org). For alternate study designs, TIDieR can be used in conjunction with the appropriate checklist for that study design (see www.equator-network.org).

Appendix E: Patient Global Impression of Change

We analyzed the patient global impression of change (PGIC) using a stratified Cochran-Mantel-Haenszel (CMH) row mean score test with table scores and one degree of freedom and found similar results to the a-priori ANOVA reported in the main manuscript. Participants in the DTxP group reported similar PGIC compared to the sham placebo group (CMH(1)=3.34, p=0.07) and better PGIC the standard care group (CMH(1)=4.65, p=0.03).

Appendix F: Subgroup analyses

Several subgroup analyses were planned a-priori for and reported here. The number of participants contributing to analyses are small, so results should be interpreted cautiously. Subgroup analyses were only conducted on the TSK measure.

We conducted the following subgroup analyses:
1. Age (< 55, ≥ 55)
2. Sex
3. Baseline TSK score (<42, ≥42)
4. Baseline NRS (<6, ≥6)
5. Baseline ODI (<35, ≥35)
6. Baseline Keele StarT Back Screening Tool (high, medium and low risk)
7. Duration of CLBP (<1, 1-5, ≥5yrs)

Fear of movement (TSK)

A planned subgroup analysis did not reveal any difference for DTx compared to sham placebo control by age (<55yr, ≥55yr), sex, TSK score at baseline (<42, ≥42), NRS at baseline (<6, ≥6), ODI (<35, ≥35) and duration of CLBP (<1, 1-5, ≥5yrs).

The subgroup analysis for Keele StarT back (high, medium or low risk) was of interest at p=0.07 because tests of interaction are generally weak and a p-value<0.1 raises interest. Compared to sham placebo control, DTxP reduced TSK at 6-8 weeks by -0.032 in the high risk group, -12.502 in the medium risk group and -2.974 in the low risk group on fear of movement. Between group differences are shown in Table F1.

Table F1. Subgroup analysis of the Keele StarT Back (high, medium, low), post-treatment.

Outcome/Group	High risk Estimate (95% CI) p-value	Medium risk Estimate (95% CI) p-value	Low risk Estimate (95% CI) p-value
Fear of pain (TSK)			
DTxP vs. Sham placebo	-0.03 (-5.88 to 5.82), p=0.99	-12.50 (-20.52 to 4.48) p<0.01	-2.97 (-16.40 to 10.45) p=0.65
DTxP vs. Standard care	-2.66 (-10.40 to 5.08) p=0.49	-9.33 (-16.39 to -2.27) p=0.01	-8.04 (-21.51 to 5.43) p=0.23
Sham placebo vs. Standard care	-2.63 (-9.94 to 4.69) p = 0.46	3.17 (-4.19 to 10.54) p=0.39	-5.07 (-20.71 to 10.58) p=0.51

Appendix G: 9 weeks follow-up results

Due to a 2-week period in which participants could complete treatment (6-8 weeks), we assessed participants at 9 weeks post-treatment as a standard assessment time-point for all participants. As can be seen from the table below, there were significant differences between the DTxP and sham placbo arms, and DTxP and standard care arms on fear of movement at 9-weeks. No other group differences were identified at 9-weeks. Please note, not all measures were assessed at week 9 to reduce assessment burden.

Table A. 9-week treatment effects from repeated measures ANCOVA (ITT sample)

Outcome/Group/Time-point	Estimate (SE)	95% CIs	P-value
Disability (ODI) post-treatment			
DTxP vs. Sham placebo	-2.18 (4.47)	-11.31 to 6.95	0.63
DTxP vs. Standard care	-5.45 (4.67)	-15.01 to 4.11	0.25
Sham placebo vs. Standard care	-3.28 (4.69)	-12.87 to 6.32	0.49
Fear of movement (TSK) post-treatment			
DTxP vs. Sham placebo	-4.46 (2.14)	-8.79 to -0.13	0.04
DTxP vs. Standard care	-5.85 (2.15)	-10.23 to -1.47	0.01
Sham placebo vs. Standard care	-1.39 (2.19)	-5.82 to 3.04	0.53
Average pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.24 (0.77)	-1.81 to 1.32	0.75
DTxP vs. Standard care	-0.39 (0.85)	-2.11 to 1.34	0.65
Sham placebo vs. Standard care	-0.14 (0.81)	-1.79 to 1.50	0.86
Maximum pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.24 (1.03)	-2.34 to 1.85	0.82
DTxP vs. Standard care	-0.39 (1.12)	-2.69 to 1.91	0.73
Sham placebo vs. Standard care	-0.15 (1.08)	-2.35 to 2.06	0.89
Minimum pain intensity (NRS) post-treatment			
DTxP vs. Sham placebo	-0.68 (0.80)	-2.32 to 0.95	0.40
DTxP vs. Standard care	-0.35 (0.89)	-2.15 to 1.46	0.70
Sham placebo vs. Standard care	0.34 (0.85)	-1.38 to 2.06	0.69

CI: confidence intervals; DTXP: Digital treatment for Pain; NRS: numerical rating scale; ODI: Oswestry Disability Index; SE: standard error; TSK: Tampa Scale for Kinesiophobia.