RESEARCH PAPER

Comparative Effectiveness of Virtual Reality Therapy, Telerehabilitation, and Conventional Physiotherapy for Disability in Chronic Low Back Pain

Bharat Bhushan[1], Monika Chawla[2], Tabish Fahim[3]
[1,3] School of Healthcare and Allied Sciences, G. D. Goenka University, Sohan- Gurugram Road, Gurugram-122103.
[2] Department of Computer Science & Engineering, Banarasidas Chandiwala Institute of Information Technology, Kalkaji, Delhi-110019, India.
Corresponding Author: Tabish Fahim (tabish.fahim@gdgu.org), Bharat Bhushan (drbbhushan@gmail.com)

Abstract
Background: Disability is the principal burden of chronic low back pain, and digitally mediated delivery may extend access to exercise-based rehabilitation, but comparative evidence holding exercise content and dose constant is lacking.
Objective: To compare virtual reality therapy, telerehabilitation, and conventional physiotherapy for reducing disability in adults with CLBP.
Methods: In this prospective, three-arm, assessor-blinded randomized trial (CTRI/2023/11/059788), 66 adults aged 25–45 years with CLBP > 12 weeks were allocated 1:1:1 by concealed permuted-block randomization. All arms received an identical 30-minute exercise protocol 5 days/week for 6 weeks. The primary outcome was change in Oswestry Disability Index (ODI%) from baseline to week 6, analyzed with a pre-specified baseline-adjusted ANCOVA (Wan, 2021, p. 1).
Results: Retention and adherence were 100%. Mean ODI% reductions were 24.91, 23.59, and 26.36 points (all $p < 0.001$; within-group $d > 5.0$), with 100% of participants achieving the ≥ 10-point MCID. The primary ANCOVA was non-significant ($F = 2.709$, $p = 0.074$); adjusted contrasts were TRT versus CPT −1.47 points (95% CI −3.54 to 0.60), VRT versus CPT +0.92 (−1.16 to 2.99), and VRT versus TRT +2.39 (0.31 to 4.47). Unadjusted ANOVA ($F = 1.939$, $p = 0.152$) and Kruskal–Wallis ($H = 3.718$, $p = 0.156$) were concordant. The non-significant result constitutes insufficient evidence of a difference rather than equivalence, which would require a pre-specified margin (Piaggio, et al., 2012; Flight & Julious, 2015); the exploratory TRT-versus-VRT contrast (−2.78 points, 95% CI −5.68 to 0.12; $d = −0.58$; $p = 0.060$) was not pre-specified, consistent with meta-analytic evidence of no significant VR benefit for disability (Opara, Kozinc, & Manojlović, 2024). No serious adverse events occurred.
Conclusions: All three delivery modes produced large, clinically meaningful 6-week disability reductions with no detected between-mode difference.
Keywords: *chronic low back pain; virtual reality therapy; telerehabilitation; conventional physiotherapy; Oswestry Disability Index; randomized controlled trial*.
How to cite this article: Bharat Bhushan, Monika Chawla, Tabish Fahim,. Comparative Effectiveness of Virtual Reality Therapy, Telerehabilitation, and Conventional Physiotherapy for Disability in Chronic Low Back Pain. Int J Drug Deliv Technol. 2026;16(83s):1-20. DOI: 10.25258/ijddt.16.83s.1

1. Introduction

Chronic low back pain (CLBP) is the leading cause of years lived with disability globally, and the disability it imposes is a principal reason patients seek care (Tagliaferri, et al., 2019, p. 8). Exercise-based physiotherapy is the standard first-line intervention, yet its effects are modest and attenuate over time, and adherence to home-based programs is poor (Slatman, Ostelo, Goor, Staal, & Knoop, 2023, p. 2); clinic-based delivery is further constrained by travel, cost, and therapist scarcity, leaving many episodes of care incomplete (Cui, et al., 2023, p. 1). Digitally mediated delivery may address these barriers: sensor-based and immersive virtual reality platforms add real-time movement tracking, biofeedback, and gamified engagement (Garofano, et al., 2025; The Effects of Virtual Immersive Gaming to Optimize Recovery (VIGOR) in Low Back Pain: A Phase II Randomized Controlled Trial), while synchronous telerehabilitation extends therapist-supervised exercise to patients who could not otherwise attend (Shi, et al., 2024; Zheng, et al., 2023, p. 6). What the literature lacks, however, is a randomized comparison of these three delivery modes within a single design holding exercise content and dose identical: pooled syntheses treat remote rehabilitation as one heterogeneous class ($I^2 > 50\%$), and the "comparator effect" — the relational context of care rather than its technological content — is routinely under-accounted in such trials (Comparison of remote rehabilitation and traditional face-to-face rehabilitation for chronic low back pain: A systematic review and meta-analysis of randomized controlled trial, p. 2).

Interpersonal context complicates delivery-mode comparisons. The therapeutic alliance — agreement on treatment goals and tasks plus a personal bond —

predicts outcomes in CLBP (Lawford, Bennell, Campbell, Kasza, & Hinman, 2020; Ferreira, et al., 2012, p. 2); because the exercise content and dose are identical across arms in this trial, any between-arm difference (or its absence) could reflect mode-dependent alliance processes rather than the exercise protocol per se (Comparison of remote rehabilitation and traditional face-to-face rehabilitation for chronic low back pain: A systematic review and meta-analysis of randomized controlled trial, p. 2).

This trial therefore compared VRT, TRT, and conventional physiotherapy for disability reduction in adults with CLBP, evaluating the effectiveness of each mode relative to the others at 6 weeks and comparing their safety and tolerability. Hypotheses were difference-detecting within a two-sided superiority framework: H_0, no between-arm difference in mean Oswestry Disability Index change from baseline to week 6; H_1, at least one arm differs. The sample (N = 66; 22 per arm) was powered for a large omnibus effect only (Cohen's f = 0.40, $\alpha = 0.05$, $1-\beta = 0.80$), so a nonsignificant result would constitute insufficient evidence of a difference — not equivalence, which requires a pre-specified margin and margin-based sample size (Prel, Hommel, Röhrig, & Blettner, 2009; Flight & Julious, 2015, p. 3). Secondary objectives were within-arm change, attainment of the ≥10-point ODI minimal clinically important difference (Vianin, 2008, p. 2), adherence, and adverse events.

2. Methods

2.1 Trial design and conduct

2.1.1 Design and Participants

This was a prospective, randomized, three-arm, assessor- and analyst-blinded trial with 1:1:1 allocation and sequential enrollment windows, reported per CONSORT 2010 (Schulz, Altman, Moher, & Group*, 2010, p. 1) and the TIDieR and TIDieR-Telehealth checklists (Hoffmann, et al., 2014; Rhon, et al., 2022). It was prospectively registered (CTRI/2023/11/059788, 14 November 2023) and approved by the GDGU Research Ethics Committee (GDGU/SoMAS/2023/168A) and the Waves Women Empowerment Trust Independent Ethics Committee (WWET/2023/IEC-AP/04). Participants were recruited from four ambulatory rehabilitation sites in the Faridabad–Gurugram National Capital Region, India (G. D. Goenka University Physiotherapy OPD; Navya Ortho Physiotherapy & Rehabilitation Centre, Faridabad; Neuromotion, Faridabad; Rehab Hub, Gurugram).

Eligible participants were adults aged 25–45 years of either sex with CLBP of more than 12 weeks' duration who provided written informed consent, including separately documented consent to randomization to a digital-mode intervention and to 6 weeks of physiotherapy (5 sessions/week) in the allocated arm. Exclusion criteria were concurrent or planned alternative back-pain treatment (surgery, nerve blocks, analgesic medication); non-spine-related or soft-tissue conditions potentially associated with back or sciatic pain; history of spinal operation, vertebral fracture, or dislocation; and severe neurological deficit (e.g., cauda equina syndrome). VRT candidates underwent additional contraindication screening (Section 2.3.4).

An *a priori* power analysis (G*Power v3.1.9.7; repeated-measures within–between ANOVA, 3 groups, 2 occasions) using a large effect size (Cohen's f = 0.40) and two-sided $\alpha = 0.05$ indicated 66 participants (22 per arm) for $1-\beta = 0.80$ (critical F = 3.143, df1 = 2, df2 = 63; achieved omnibus power 0.818); 66 were recruited. The arms were enrolled in sequential, non-overlapping windows — CPT (20 November 2023–4 January 2024), then TRT (16 April–15 June 2024), then VRT (17 October–5 December 2024) — reflecting the staged availability of each mode's infrastructure; allocation within each window was concealed and drawn from the single pre-generated permuted-block sequence (Section 2.1.2), with all four sites contributing in every window. Enrollment window and treatment arm are therefore perfectly confounded, so calendar-time effects cannot be separated from delivery-mode effects; this is acknowledged in the Limitations (Section 4.1).

The trial was powered for the difference-detecting omnibus comparison only, not for equivalence: margin-based designs of comparable CLBP trials have required >100 participants (Cui, et al., 2023; Zheng, et al., 2023, p. 7), so non-significant between-arm findings must be read as a lack of evidence of a difference rather than as equivalence (Sections 3.1.3 and 4.1).

2.1.2 Randomization and Blinding

Participants were allocated 1:1:1 to VRT, TRT, or CPT by computer-generated permuted-block randomization (fixed block size of 6). The allocation sequence was generated using the Research Randomizer tool by a statistician independent of recruitment and outcome assessment before enrollment began. Allocation concealment was maintained with sequentially numbered, opaque, sealed envelopes prepared in advance and held by a researcher independent of enrollment and assessment until assignment; two co-investigators assigned each enrolled participant by opening the next envelope, and assignment followed the same pre-generated sequence regardless of site. Blinding of participants and treating therapists to the visibly and experientially distinct delivery modes was not feasible (Cashin, et al., 2020, p. 12); the outcome assessor and the statistician conducting the analysis were blinded throughout, yielding an assessor- and analyst-blinded design.

2.1.3 Statistical Analysis

The pre-specified primary analysis was a baseline-adjusted analysis of covariance with week-6 ODI% as the dependent variable, baseline ODI% as the covariate, and treatment arm as the fixed factor,

estimated by ordinary least squares on the intention-to-treat population (Wan, 2021, p. 1); this parameterization is algebraically identical to the change-score ANCOVA originally pre-specified, differing only in the covariate coefficient, with identical arm contrasts, confidence intervals, and F statistics (Clifton & Clifton, 2019, p. 2). Covariate coefficients are reported on the improvement scale (dependent variable = baseline ODI% − week-6 ODI%); on the equivalent week-6 post-score scale each coefficient shifts by exactly 1 — here $\beta = 1.062$ on the improvement scale corresponds to ≈ -0.06 (1 − 1.062) on the post-score scale — with no change to contrasts, confidence intervals, or test statistics (Clifton & Clifton, 2019, p. 2). Baseline adjustment was specified a priori because ANCOVA yields unbiased treatment-effect estimates under baseline imbalance and is more precise than unadjusted or change-score analyses (Egbewale, Lewis, & Sim, 2014, p. 1). Assumption checks comprised the arm-by-baseline interaction (homogeneity of slopes), residual Q–Q and residuals-versus-fitted plots, and Shapiro–Wilk and Levene tests; given the limited power of such tests with 22 participants per arm, these were interpreted diagnostically alongside graphics, with no conditional selection of the analytical method (Shatz, 2023; Zygmont, 2023, p. 4). Pre-specified robustness analyses were the unadjusted one-way ANOVA and the Kruskal–Wallis test, with parametric, non-parametric, and variance-robust results reported together (Zygmont, 2023, p. 11). Within-group change was assessed by paired t-test corroborated by the Wilcoxon signed-rank test; categorical outcomes by chi-square or Fisher's exact test as appropriate. Missing data were to be handled by multiple imputation under a missing-at-random assumption with missing-not-at-random sensitivity analysis; this was not required because the analyzed dataset contained no missing observations. Significance was set at $P < 0.05$ (two-tailed); analyses were conducted in Python (SciPy 1.17), reporting test statistics, exact P values, effect sizes (η^2, Cohen's d), and 95% confidence intervals, with adjusted between-arm differences additionally interpreted against the pre-specified ≥10-point ODI minimal clinically important difference (Vianin, 2008, p. 2).

2.1.4 Treatment Protocols

All three arms received an identical structured spinal exercise protocol at a standardized dose — once daily, 30 minutes per session, five days per week, for six weeks (30 sessions) — comprising a 10-minute warm-up of slow body movements; core-strengthening and flexion exercises (sit-ups, knee-to-chest, toe-touch, squat training); extension exercises (prone press-up, back-arch variations); and a 5-minute cool-down stretch. The arms differed only in delivery mode, with full itemized reporting against the TIDieR and TIDieR-Telehealth checklists provided in Table 4 (Hoffmann, et al., 2014; Rhon, et al., 2022).

In the two digital arms, all sessions were delivered on-site at the study facility, with the participant in one treatment room and the therapist in an adjacent room connected via the remote link — that is, therapist–participant separation between rooms rather than home-based delivery — to secure attendance and adherence, consistent with evidence that unsupervised, self-reported exercise does not reliably reproduce supervised treatment effects (Zheng, et al., 2023, p. 2) and that remote rehabilitation delivery should include verification of a safe, appropriate environment (Baroni, et al., 2023, p. 1).

Virtual Reality Therapy. The identical protocol was delivered within an immersive virtual environment via a Utopia 360 head-mounted display. The participant performed the protocol in a space- and safety-checked treatment room, viewing a 360° first-person exercise demonstration inside the headset, while a registered physiotherapist in an adjacent monitoring room observed and guided the session in real time through a remote-monitoring link, with correction and cueing delivered by audio (and, where available, video) rather than physical contact. VRT participants underwent additional pre-randomization contraindication screening (Section 2.3.4) and the arm-specific safety protocol covering cybersickness, falls risk, and headset infection control (Sections 2.3.5–2.3.7). VRT was activated last (17 October–5 December 2024); fidelity was tracked via clinician self-report checklists, with technology-specific deviations (headset malfunction, connectivity loss, cybersickness-related early termination) logged.

Telerehabilitation Therapy. The identical protocol was delivered synchronously via Microsoft Teams, with the participant using their own smartphone, tablet, or laptop. The physiotherapist observed the webcam feed in real time, cued exercises verbally, and corrected technique on the basis of visual assessment alone, without physical palpation; exercise tailoring therefore relied on participant self-report combined with remote visual monitoring rather than the direct in-person assessment available in CPT. TRT participants completed the same pre-randomization digital-literacy screening and onboarding pathway as VRT (Section 2.3.1), plus TRT-specific safety overlays: a video-on verification of the environment, lighting, and setup at the start of each session, a study diary with photo/video documentation of session completion, and a technical helpline for real-time troubleshooting. TRT was activated in the second enrollment window (16 April–15 June 2024); fidelity was tracked via session-recording review or clinician self-report checklist, with connectivity-related deviations (call dropout, platform failure, rescheduling) logged.

Conventional Physiotherapy. The identical protocol was delivered in the personal mode — participant and registered physiotherapist together in the same treatment room — under continuous, one-to-one supervision for the full duration of every session, using standard clinic treatment space, mats, and equipment. Education was delivered through standardized printed take-home handouts, with the therapist demonstrating each exercise before participant performance and reinforcing technique verbally. Tailoring and progression were based on direct in-person assessment, including hands-on cueing, physical correction of technique, and manual palpation. CPT was activated first (20 November 2023–4 January 2024); fidelity was monitored by a senior physiotherapist using in-person checklist observation (Hildebrand, et al., 2012, p. 1).

Beyond these procedural standardizations, no validated process measure of interaction quality (e.g., the Working Alliance Inventory or a presence/engagement scale) was administered in any arm; mode-dependent differences in therapeutic alliance therefore cannot be quantified and are addressed as a limitation (Section 4.1).

2.2 Outcome Assessment

2.2.1 Outcome Measures

The primary outcome was functional disability on the Oswestry Disability Index (ODI, 0–100%. higher = greater disability), a valid, reliable, and responsive instrument for CLBP (Vianin, 2008, p. 1), administered by a blinded assessor at baseline and week 6, with no interim primary-outcome assessment. Secondary outcomes were achievement of the pre-specified ≥10-point ODI minimal clinically important difference (Vianin, 2008, p. 2);

2.2.2 Adherence and Retention

Adherence (sessions completed ÷ prescribed) and dropout rates were recorded for every participant in all arms, as was adverse event reporting.

2.2.3 Adverse Events and Safety Outcomes

Adverse events per ICH-GCP definitions, classified on severity, relatedness, and expectancy axes (Dennett, et al., 2022, p. 6) and adjudicated by an independent, allocation-blinded clinician on a five-point causality scale (Koh, et al., 2015, p. 12). Events of specific interest were falls during exercise, exercise-attributable musculoskeletal flare, technology-related events (connectivity failure, audiovisual disturbance, privacy breach) in TRT/VRT, and cybersickness in VRT, monitored with the Fast Motion Sickness Scale, the Simulator Sickness Questionnaire, and pre-specified stop criteria (Section 2.3.5); a further digital-arm outcome was the proportion requiring Tier A/B accommodation or Tier C conversion, logged with structured verbatim reasons (Section 2.3.2). All assessments fell within the 6-week intervention — no post-intervention follow-up was scheduled, so the durability of improvements beyond week 6 cannot be determined — while adherence and adverse-event surveillance ran continuously across all 30 sessions per participant.

2.3 Safety Protocol for the Digital-Care Arms

Sections 2.3.1–2.3.3 were common to both digital arms; Sections 2.3.4–2.3.7 were VRT-specific; CPT relied on continuous in-person supervision alone.

2.3.1 Digital Literacy Screening and Onboarding

Before randomization, TRT/VRT candidates completed a 5-item Brief Digital Literacy Checklist (device/connectivity access, unassisted video-call answering, guided app installation, home internet ≥10 Mbps, caregiver availability) and the 8-item eHealth Literacy Scale (Norman & Skinner, 2006, p. 1). Low scores did not exclude but routed participants to enhanced onboarding (a 20–30-minute in-person session, illustrated quick-reference guide, caregiver support person, and technical helpline), with digital confidence reassessed at sessions 1 and 2; as a self-report measure validated in a youth web-based context, the eHEALS served as a general literacy flag rather than a predictor of headset-specific proficiency (Norman & Skinner, 2006, p. 1).

2.3.2 Tiered Accommodation and Rescue Pathway

A predefined three-tier pathway retained participants unable to operate the technology: Tier A accommodation (facilitated onboarding, study-provided tablet/SIM, pictorial exercise cards), Tier B caregiver-mediated delivery, and Tier C rescue conversion to CPT, with the original allocation preserved for intention-to-treat analysis and the conversion logged as a protocol deviation; no tier was activated in this trial, so the ITT and per-protocol populations were identical.

2.3.3 Adverse Events Definition, Reporting and Escalation

Adverse events followed ICH-GCP definitions as any untoward medical occurrence regardless of causality (Moussa, et al., 2025; Koh, et al., 2015, p. 2); serious events comprised death, life-threatening illness, hospitalization, persistent disability, or congenital anomaly (Moussa, et al., 2025, p. 2). Expected events included cybersickness symptoms (nausea, dizziness, disorientation, headache, eye strain) (Costa, et al., 2025, p. 19), headset-related discomfort (Birckhead, et al., 2018, p. 5), and privacy- or data-related incidents (Yau, et al., 2024, p. 2). Events were captured continuously (direct observation, app-based diary (Daly, Gianoudis, Hall, Mundell, & Maddison, 2020, p. 3), weekly check-ins, session metadata) and reported per the CONSORT Harms extension (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials, 2023): SAEs to the sponsor and the Data Safety Monitoring Board within 24 hours and the ethics committee within 7 days (Shaheed, Maher, Furmage, Hoffmann, & McLachlan, 2022; Wood, et al., 2024), with SAEs, VR-related events, exercise-stopping events, and

events occurring in ≥5% of either arm reported individually (Costa, et al., 2025, p. 19); participants retained the right to terminate any session at any time (Costa, et al., 2025, p. 19). Telerehabilitation events are rare (0.31%; 3.1 per 1,000 sessions) and are captured only when deliberately measured (Shnitzer, et al., 2024). Where visual assessment was judged insufficient, a three-stage response governed the digital arms (Alrwaily, 2025): modification (exercise regression, reduced range/load, or self-palpation instruction (Russell, Truter, Blumke, & Richardson, 2010; Baroni, et al., 2023)), pause (termination on symptom escalation, suspected red-flag features, or two unsuccessful correction attempts), and referral (reassessment by a senior physiotherapist within 48 hours, Tier C transition where continued remote delivery was judged unsafe, or medical review), with each activation logged as a protocol deviation.

2.3.4 VRT-Specific Exclusion Screening

VRT-specific safety began with additional exclusion screening — vestibular/oculomotor disorders or severe cybersickness history, photosensitive epilepsy, recent traumatic brain injury, uncontrolled migraine, corrected acuity worse than 20/200, uncontrolled hypertension or myocardial infarction within 3 months, MMSE < 24, active psychosis, severe anxiety, or substance use disorder, and facial wounds, dermatitis, or silicone allergy — and a supervised 5-minute headset familiarization.

2.3.5 Cybersickness Monitoring and Stop Criteria

Cybersickness was monitored with the FMS (0–20) at 10-minute intervals and the SSQ after every immersion (Keshavarz & Hecht, 2011; Kennedy, Lane, Berbaum, & Lilienthal, 1993); continuous immersion was capped at 20 minutes with a mandatory ≥5-minute seated rest, and pre-specified stop criteria (weighted SSQ ≥ 40, FMS ≥ 7, participant request, marked pallor, frank disequilibrium) triggered termination and review, two consecutive symptom-triggered stoppages mandating a headset-refitting review and possible downgrade to non-immersive VR or TRT; no VRT session terminated early (30/30 sessions, 100%).

2.3.6 Fall Risk and Environmental Safety

Fall risk was screened at baseline with the CDC STEADI battery (Timed Up & Go, 30-Second Chair Stand, Four-Stage Balance) and a fall-history interview (Lusardi, et al., 2016; Palumbo, Palmerini, Bandinelli, & Chiari, 2015; Pillay, et al., 2024, p. 33). All sessions were seated for the first two weeks; standing progression required two consecutive tolerance-checked sessions (SSQ < 40, FMS < 7) and a repeat TUG ≤ 12 s, adapted from older-adult thresholds (Bischoff, 2003; Schoene, et al., 2013), with participants failing the gait remaining seated throughout. Mitigation combined a hazard-free room with the boundary system enabled (Visual feedback and guided balance training in an immersive virtual reality environment for lower extremity rehabilitation, p. 5), a stable chair within reach (Stamm, Dahms, & Müller-Werdan, 2020, p. 9), a spotter during all standing tasks, and therapist authority to terminate sessions (Costa, et al., 2025, p. 19); falls were recorded as adverse events of specific interest.

2.3.7 Headset Infection Control and Hygiene

Headset decontamination followed a standard operating procedure (disposable face cover, post-use disinfection of contact surfaces, documented Cleaning Log with a monthly independent audit of 10% of events). Where visual assessment was judged insufficient, a three-stage response governed the digital arms (Alrwaily, 2025): modification (exercise regression, reduced range/load, or self-palpation instruction (Russell, Truter, Blumke, & Richardson, 2010; Baroni, et al., 2023)), pause (termination on symptom escalation, suspected red-flag features, or two unsuccessful correction attempts), and referral (reassessment by a senior physiotherapist within 48 hours, Tier C transition where continued remote delivery was judged unsafe, or medical review), each activation logged as a protocol deviation.

2.4 Ethical Considerations

Written informed consent was obtained from all participants, including separate consent to randomization to a digital-mode intervention and its digital-care safety screening; participants retained the right to withdraw without prejudice and, if unable to continue a digital arm, were transitioned to CPT under the Tier C rescue pathway rather than withdrawn. Video/telehealth session data and case report forms were stored on access-controlled institutional systems; any privacy breach was treated as an adverse event of special interest.

3. Results

3.1 Trial Results

3.1.1 Baseline Comparability

Baseline ODI% was comparable across arms (CPT 34.13 ± 2.73; TRT 34.27 ± 3.32; VRT 34.64 ± 2.98; one-way ANOVA F = 0.164, p = 0.849; Table 1).

Table 1. Baseline and Week-6 ODI% by Arm: Descriptive Statistics, Within-Group Change, and Unadjusted Between-Arm Comparisons

Panel A — Descriptive statistics and group-specific means *with* 95% CIs

Arm	n	Baseline ODI% (*M* ± SD; 95% CI)	Week 6 ODI% (*M* ± SD; 95% CI)	Mean change (95% CI)	Paired *t* (df = 21)	*p*	Cohen's *d*
CPT	22	34.13 ± 2.73 (32.92–35.34)	9.22 ± 3.22 (7.79	24.91 (22.91 – 26.91)	25.89	<0.001	5.52

			–10.65)				
TRT	22	34.27 ± 3.32 (32.80–35.74)	10.69 ± 3.50 (9.14 – 12.24)	23.59 (21.67 – 25.51)	25.50	<0.001	5.44
VRT	22	34.64 ± 2.98 (33.32–35.96)	8.27 ± 3.51 (6.71 – 9.83)	26.36 (24.08 – 28.64)	24.01	<0.001	5.12

Table note: 95% CIs for the means were computed from the reported within-arm means and SDs using the *t*-distribution ($t = 2.080$, df = 21).

Panel B — Between-group comparisons of ODI% change

Unadjusted (independent-samples t-test):

Comparison	Mean difference (points) [a]	95% CI	*t* (df = 42)	*P* (unadjusted)	P_b	Cohen's *d*
CPT vs TRT	1.32	−1.37 to 4.00	0.992	0.327	0.981	0.30
CPT vs VRT	−1.45	−4.39 to 1.49	−0.996	0.325	0.974	−0.30
TRT vs VRT	−2.78	−5.68 to 0.12	−1.935	0.060	0.179	−0.58

Table notes: [a] Positive difference favours the first-named arm. [b] Bonferroni threshold α = 0.0167 per comparison; no contrast was significant. 95% CIs for the means were computed from the reported within-arm means and SDs (t = 2.080, df = 21).

Pairwise baseline differences were small — TRT vs CPT 0.14 points (95% CI −1.71 to 1.99), VRT vs CPT 0.51 (−1.23 to 2.25), VRT vs TRT 0.37 (−1.55 to 2.29) — all intervals including zero and lying well below the 10-point MCID; the pre-specified analysis nevertheless adjusted for baseline ODI% (Egbewale, Lewis, & Sim, 2014, p. 1).

3.1.2 Assumption Checks

Shapiro–Wilk tests did not reject normality in any arm (CPT W = 0.978, p = 0.888; TRT W = 0.951, p = 0.327; VRT W = 0.954, p = 0.373); Levene's test was non-significant (F = 0.052, p = 0.949); and the arm-by-baseline interaction was non-significant (F = 1.21, p = 0.305), supporting the common-slope model. ANCOVA residuals tracked the normal quantiles (Shapiro–Wilk W = 0.986, p = 0.668) without heteroscedasticity (p = 0.424). Given 22 participants per arm, formal tests were interpreted diagnostically alongside the graphical checks (Shatz, 2023; Zygmont, 2023, p. 4), and the pre-specified robustness analyses corroborated every between-group inference.

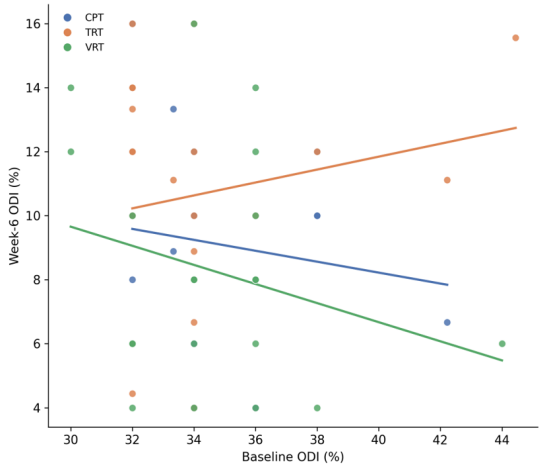

Figure 1. Week-6 Oswestry Disability Index (ODI) percentage versus baseline ODI percentage, by treatment arm, with arm-specific fitted linear regression lines.

Note. N = 66 (22 per arm). Lines are simple linear regressions of week-6 ODI% on baseline ODI% fitted separately within each arm (CPT: slope = -0.17, r = -0.14, p = 0.522; TRT: slope = 0.20, r = 0.19, p = 0.393; VRT: slope = -0.30, r = -0.25, p = 0.254); none of the individual within-arm slopes differed significantly from zero. The formal test of whether these slopes differ from one another — the arm-by-baseline interaction term in the extended ANCOVA model — was likewise non-significant, F(2, 60) = 1.21, p = 0.305, consistent with (though not conclusive proof of) a common baseline–outcome relationship across arms. Because each arm's baseline ODI% spans a narrow range (30%–44%) and each regression is estimated from only 22 points, the individual-arm slope estimates are imprecise and should be interpreted alongside the formal interaction test rather than in place of it. CPT = conventional physiotherapy; TRT = telerehabilitation therapy; VRT = virtual reality therapy.

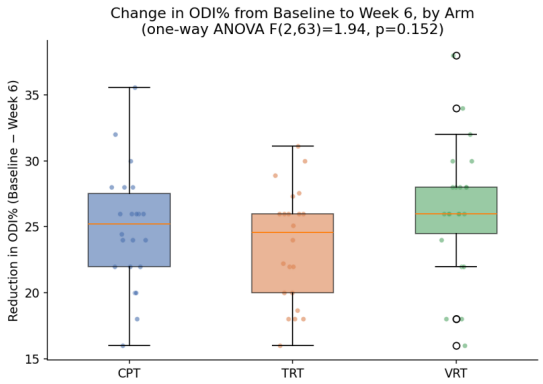

Figure 2. Change in ODI% (Baseline − Week 6) by arm. Boxes show IQR; dots are individual participants; circles are outliers.

Graphical checks. Normal Q–Q plots of within-arm change scores and of the residuals of the pre-specified ANCOVA model were inspected in conjunction with the formal tests, following the recommendation that the normality assumption be evaluated graphically rather than by significance test alone (Zygmont, 2023, p. 305); the plotted points tracked the theoretical quantiles with no systematic departure at the tails, and the residuals-versus-fitted plot showed no evidence of heteroscedasticity. Taken together, the parametric analyses were retained as appropriate: the diagnostics support (but do not prove) the assumptions, and the concordant Kruskal–Wallis and Welch results provide distribution-free and variance-robust corroboration of every between-group inference.

3.1.3 Primary Outcome Analysis

The pre-specified baseline-adjusted ANCOVA (covariate $\beta = 1.062$ on the improvement scale, i.e., change = baseline minus week-6 ODI%; equivalent to ≈ -0.06 on the week-6 post-score scale; $p < 0.001$; Section 2.1.3) found no significant arm effect ($F = 2.709$, $p = 0.074$; Table 2).

Table 2. Baseline-Adjusted Between-Group Differences in ODI% Change from Baseline to Week 6

Contrast	Adjusted difference (points)	95% CI	P
TRT vs CPT	−1.47	−3.54 to 0.60	0.160
VRT vs CPT	+0.92	−1.16 to 2.99	0.380
VRT vs TRT	+2.39	0.31 to 4.47	0.025

Model notes: ANCOVA with baseline ODI% as covariate (improvement scale, dependent variable = baseline − week-6 ODI%; $\beta = 1.062$, $P < 0.001$; on the week-6 post-score scale the coefficient is −0.06); omnibus $F = 2.709$, $P = 0.074$ (pre-specified primary analysis); CPT = reference category; negative differences indicate less improvement than CPT. ANCOVA is the appropriate primary estimator for this pre–post design (Wan, 2021, p. 15). Omnibus tests: ANCOVA $F = 2.709$, $P = 0.074$ (pre-specified primary); one-way ANOVA $F = 1.939$, $P = 0.152$, $\eta^2 = 0.058$; Kruskal–Wallis $H = 3.718$, $P = 0.156$.

Equivalence note: No equivalence or non-inferiority margin was pre-specified; the sample size was powered for the omnibus difference-detecting comparison (Cohen's $f = 0.40$) rather than for a margin, and no one-sided margin-based analysis was performed; the non-significant omnibus tests therefore constitute a lack of evidence of a difference, not evidence of equivalence (Piaggio, et al., 2012; Flight & Julious, 2015). Descriptively, all pairwise 95% CIs fall within ±10 ODI points — the pre-specified MCID threshold (Vianin, 2008, p. 2) — and an exploratory post hoc TOST assessment (90% CIs: −0.92 to 3.56; −3.90 to 1.00; −5.20 to −0.36; all $p < 0.001$) was concordant with between-arm differences within that range (Lauzon & Caffo, 2009, p. 3). Because the margin was anchored to the MCID post hoc rather than fixed prospectively and independently of the data (Linde & Ravenzwaaij, 2023, p. 4), this reading is hypothesis-generating and consistent with equivalence within the MCID — not a demonstration of it.

Adjusted contrasts with CPT as reference were TRT vs CPT −1.47 points (95% CI −3.54 to 0.60; $p = 0.160$), VRT vs CPT +0.92 (−1.16 to 2.99; $p = 0.380$), and VRT vs TRT +2.39 (0.31 to 4.47; $p = 0.025$) — the last nominally significant but not pre-specified and non-significant after Bonferroni adjustment ($\alpha = 0.0167$). Sensitivity analyses were concordant: unadjusted one-way ANOVA $F = 1.939$, $p = 0.152$, $\eta^2 = 0.058$; Kruskal–Wallis $H = 3.718$, $p = 0.156$; Welch's ANOVA $F(2, 41.8) = 1.84$, $p = 0.171$. Unadjusted pairwise contrasts were CPT vs TRT +1.32 points (−1.37 to 4.00; $p = 0.327$), CPT vs VRT −1.45 (−4.39 to 1.49; $p = 0.325$), and TRT vs VRT −2.78 (−5.68 to 0.12; $p = 0.060$; $d = -0.58$).

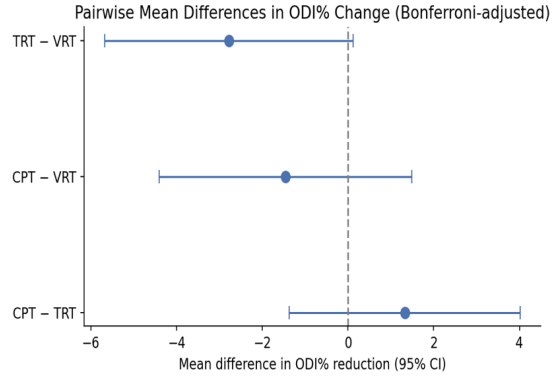

Figure 3. Pairwise mean differences in ODI% reduction with 95% CI (unadjusted t-tests, Bonferroni-corrected significance threshold).

All between-arm 95% CIs lay within ±10 ODI points, compatible with at most a modest difference

but too wide to exclude one; the result therefore constitutes insufficient evidence of a difference rather than equivalence, which would require a pre-specified margin and margin-based power (Piaggio, et al., 2012; Flight & Julious, 2015). An exploratory post hoc TOST with the interval anchored to the ±10-point MCID (Vianin, 2008, p. 2) found all pairwise 90% CIs within the margin (all $p < 0.001$); because the margin was post hoc, this is hypothesis-generating only (Lauzon & Caffo, 2009; Linde & Ravenzwaaij, 2023, p. 3).

3.1.4 Within-Group Change

All arms showed large, significant ODI% reductions: mean changes of 24.91 points (95% CI 22.91–26.91) in CPT, 23.59 (21.67–25.51) in TRT, and 26.36 (24.08–28.64) in VRT (paired $t = 25.89$, 25.50, and 24.01 respectively; all $p < 0.001$; $d = 5.52$, 5.44, 5.12), corroborated by Wilcoxon signed-rank tests; individual reductions ranged 16.0–35.6, 16.0–31.1, and 16.0–38.0 points.

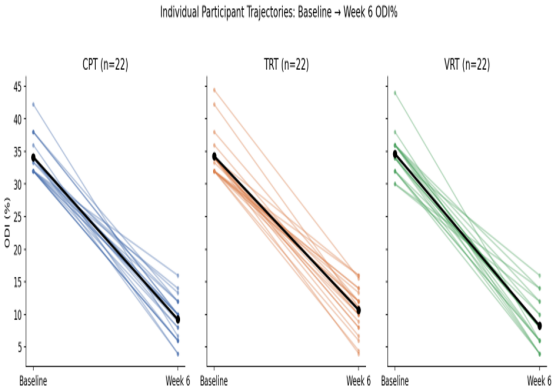

Figure 4. Individual participant trajectories from baseline to Week 6 by arm (black line = group mean). The trajectories were uniformly steep and consistently downward across all three arms.

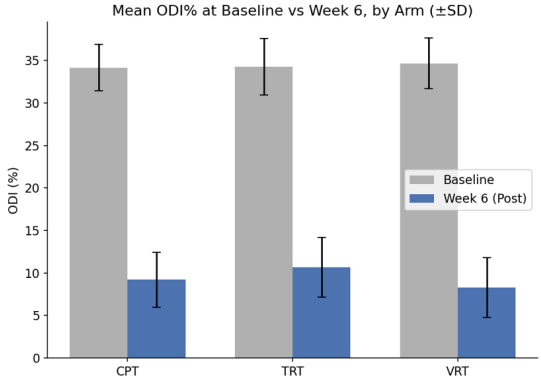

Figure 5. Mean ODI% at baseline and Week 6 by arm (± SD).

3.1.5 Minimal Clinically Important Difference

All of participants (22/22) in each arm met the pre-specified ≥10-point MCID (Vianin, 2008, p. 2); with no variability across outcome categories the chi-square test is degenerate ($\chi^2 = 0$, $p = 1.0$) and uninterpretable, so attainment is reported descriptively. The floor compression is noteworthy: the smallest individual reduction (16.0 points) exceeded the threshold by 60%, and Week 6 means (8.3–10.7 points) **fell within** the ≤20% "no-disability" region (Schwind, Learman, O'Halloran, Showalter, & Cook, 2013, p. 2), limiting MCID's discriminative value in this cohort.

3.1.6 Distributional Check

Change-score distributions were unimodal and broadly overlapping across arms; diagnostic plots of the primary ANCOVA model corroborated the formal checks.

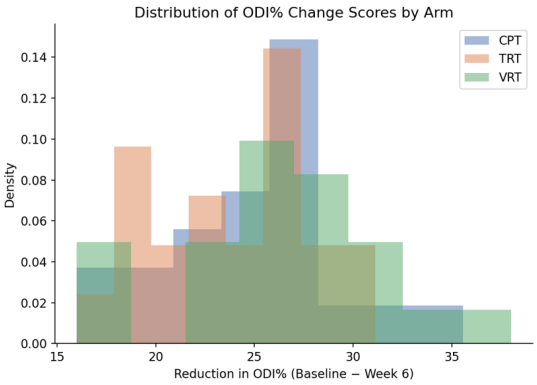

Figure 6. ODI% change scores by arm, showing broadly overlapping, unimodal, and roughly symmetric distributions consistent with the normality tests above.

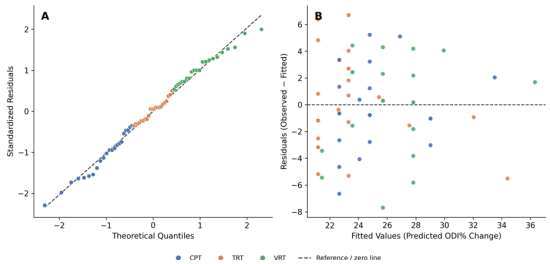

Figure 7. Diagnostic plots for the primary baseline-adjusted analysis of covariance (ANCOVA) model. (A) Normal quantile–quantile (Q–Q) plot of standardized residuals. (B) Residuals plotted against fitted values.

Note. The ANCOVA model regressed the change in Oswestry Disability Index (ODI) percentage score (baseline minus week 6) on treatment arm and baseline ODI% as a covariate ($N = 66$). In panel A, standardized residuals are plotted against the theoretical quantiles of a standard normal distribution, with the dashed line showing the expected pattern under normality; the close alignment of points to this line indicates no material departure from normality (Shapiro–Wilk $W = 0.986$, $p = .668$). In panel B, residuals (observed minus fitted) are plotted against the model's fitted values, with the dashed line marking zero residual; the even, structureless scatter with no funnel or fan pattern indicates homoscedasticity, confirmed by a formal test regressing squared residuals on fitted values ($b = -0.39$, $p = .424$). Together, panels A and B support the normality and homoscedasticity assumptions underlying the primary ANCOVA model. CPT = conventional physiotherapy; TRT =

telerehabilitation therapy; VRT = virtual reality therapy.

3.1.7 Synthesis

The trial did not detect a statistically significant between-arm difference in disability at 6 weeks: H_0 was not rejected by the primary ANCOVA ($F = 2.709$, $p = .074$) or by any sensitivity analysis. The observed omnibus effect ($\eta^2 = 0.058$) fell below the large effect assumed in the power calculation, and margin-based designs in comparable CLBP trials have required 74–120 participants (Cui, et al., 2023; Zheng, et al., 2023, p. 7), so the nonsignificant result is inconclusive rather than negative (Smith, et al., 2020; Manchikanti, 2008). The exploratory TRT vs VRT contrast (−2.78 points; 95% CI −5.68 to 0.12; $d = −0.58$; $p = .060$) was not pre-specified, did not approach the Bonferroni threshold, and is best read as an imprecise numerical difference — consistent with meta-analytic evidence finding no significant VR benefit for disability (Opara, Kozinc, & Manojlović, 2024). Taken with the uniform within-group improvement ($d > 5.0$ in all arms), the pattern is consistent with the shared exercise protocol — rather than delivery mode — as the principal driver of recovery, and with the wider literature showing digitally mediated delivery can match in-person outcomes when content and dose are held constant (Cui, et al., 2023, p. 1).

3.2 Adherence, Retention, and Protocol Deviations

Adherence and retention were complete in every arm. All 66 randomized participants (22 per arm) completed all 30 prescribed sessions and the week-6 outcome assessment — a 100% retention rate with no withdrawals — so the intention-to-treat and per-protocol populations were identical (Section 2.1.3). Complete retention also obviated any need for multiple imputation: because no outcome observations were missing, the pre-specified missing-data procedure (Section 2.1.3) was never invoked, and the analyzed dataset contained all randomized participants (Cashin, et al., 2020, p. 12). Session-level adherence (sessions completed divided by sessions prescribed; Section 2.2.2) was 30/30 sessions (100%) in CPT, TRT, and VRT alike, with every session completed as planned and no missed-session reasons logged. The absence of missing data follows from the same fact: with adherence and retention at 100%, there were no half-completed or unattended sessions from which observations could be lost.

This level of session completion is high relative to the wider evidence base in both delivery contexts. In sensor-based remote-rehabilitation and VR programs for chronic low back pain, reported completion rates are typically 73%–90%, and gamification or real-time feedback is usually required to sustain engagement (Garofano, et al., 2025); in conventional physiotherapy, compliance with therapist-prescribed advice and home exercises is characteristically low, largely attributable to lack of patient motivation (Slatman, Ostelo, Goor, Staal, & Knoop, 2023, p. 2). The present 100% completion across all three arms therefore supports the feasibility of therapist-attended digital delivery (Section 2.1.4), although the on-site, therapist-supervised design means the observed adherence reflects secured rather than home-based engagement (Section 4.1).

No digital-mode accommodation or rescue pathway was activated: no participant received Tier A supported onboarding, Tier B caregiver-mediated delivery, or Tier C conversion to CPT (Section 2.3.2), and no crossovers occurred. Consequently, the deviation log for the digital arms was empty, and no VRT session was terminated early against the pre-specified cybersickness stop criteria (30/30 sessions; Section 2.3.5). Treatment fidelity was monitored by clinician self-report checklists and — for CPT — structured in-person checklist observation by a senior physiotherapist, methods consistent with the recommended approach of demonstrating therapist adherence and competence against the protocol (Hildebrand, et al., 2012, p. 1). The fidelity record was correspondingly complete: zero technology-related deviations (connectivity failure, platform failure, rescheduling) were logged in TRT and zero headset-, connectivity-, or cybersickness-related deviations in VRT.

Two documentation gaps bound these results. First, screening-failure counts by exclusion criterion were not recorded, so differential selection across the sequential enrollment windows cannot be quantified (Section 4.1). Second, the aggregate adherence metric was not decomposed by week; with every session completed in every arm, the session-level data carry no temporal variation from which engagement trajectories could be modelled. Both gaps temper, but do not alter, the descriptive conclusion: delivery-mode-independent adherence and retention of 100% across 30 sessions per participant. Adherence, retention, and delivery integrity outcomes are summarized in Table 3; adverse events are reported in Section 3.4.

Table 3. Adherence, Retention, and Delivery Integrity Outcomes by Arm

Outcome	CPT	TRT	VRT
Randomized (intention-to-treat), 22	22	22	22
Completed week 6 assessment, 22 (100%)	22	22	22
Sessions completed as planned, 22 (100%)	30/30	30/30	30/30
Missed sessions, 22	0	0	0

Technology-related deviations logged, 22	—	0	0[a]
Early-terminated sessions, 22	—	—	0[b]
Tier A/B/C activations, 22	—	0	0
Crossovers, 22	—	0	0

Table note: Dash (—) indicates not applicable to that arm. [a] Including headset-, connectivity-, or cybersickness-related deviations. [b] The pre-specified cybersickness stop criteria (weighted SSQ ≥ 40; FMS ≥ 7) were not met in any VRT session. Fidelity monitoring procedures are detailed in Section 2.1.4 and summarized per the TIDieR checklist in Table 4.

3.3 Safety and Tolerability Monitoring

Safety and tolerability were monitored continuously across all 30 sessions per participant in every arm (Section 2.2.3), and the complete-session record reported in Section 3.2 (30/30 sessions in each arm; zero technology-related deviations) already establishes the outer safety envelope. This section reports the modality-specific monitoring outcomes that the deviation log summarizes by count.

Cybersickness monitoring in VRT. Cybersickness — the principal VR-specific tolerability concern (Birckhead, et al., 2018, p. 5) — was monitored with the Fast Motion Sickness Scale (0–20) at 10-minute intervals during immersion and the Simulator Sickness Questionnaire after every immersion, against pre-specified stop criteria (weighted SSQ ≥ 40; FMS ≥ 7; participant request; marked pallor; frank disequilibrium; Section 2.3.5). No VRT session met any criterion: all 30 sessions per participant ran to completion with no early termination and no participant-requested stoppage, and the two-consecutive-symptomatic-stoppage rule that would have mandated a headset-refitting review or downgrade to non-immersive VR or TRT was never engaged. Continuous immersion was capped at 20 minutes with a mandatory ≥ 5-minute seated rest — a ceiling consistent with evidence that the risk of vertigo increases notably beyond roughly 30 minutes of immersive use (Slatman, Ostelo, Goor, Staal, & Knoop, 2023, p. 4) — and the VRT deviation record (Section 3.2) contains no indication of a cap or rest-rule violation. Individual FMS/SSQ values below the stop thresholds are not presented here: the protocol's decision rules were threshold-based; no threshold was reached, and sub-threshold symptom burden therefore cannot be quantified from the recorded outcomes. The reported outcome is the absence of threshold-level cybersickness across all 660 VRT sessions (22 participants × 30 sessions). This pattern is consistent with the wider VR-in-CLBP evidence, in which adverse events are predominantly mild and self-limited, with cybersickness the most commonly reported (Spiegel, et al., 2026), and the symptom domains monitored here — nausea, dizziness, disorientation, headache, eye strain — match the standard cybersickness terminology recommended for VR trials (Costa, et al., 2025, p. 19). The rapid-stop protections, in which the supervising therapist held authority to terminate a session immediately and participants retained the unqualified right to stop at any time (Costa, et al., 2025, p. 19), were never exercised.

Falls and environmental safety. Falls were pre-specified events of special interest, and none occurred in any arm. In immersive headset-based exercise, the inability to perceive the surrounding room is the principal falls hazard, which is why home-delivered VR programmes require a proven safety concept and explicit emergency procedures (Stamm, Dahms, & Müller-Werdan, 2020, p. 9). That hazard context did not apply here: VRT sessions were delivered on-site in a space- and safety-checked room with the boundary system enabled, a stable chair within reach, and a spotter present during all standing tasks — the same environmental mitigations reported in supervised VR exercise research, in which no participant fell (Visual feedback and guided balance training in an immersive virtual reality environment for lower extremity rehabilitation, p. 5). Standing-phase progression was governed by the pre-specified tolerance gate (two consecutive sessions with SSQ < 40 and FMS < 7 plus a repeat Timed Up and Go ≤ 12 s; Section 2.3.5); because every session completed as planned, no gate-triggered termination, suspension, or fall occurred that would require separate reporting.

Digital-process safety in TRT. Each TRT session opened with a video-on verification of the environment, lighting, and setup, with completion documented in a study diary with photo/video evidence and a technical helpline available for real-time troubleshooting (Section 2.1.4). No connectivity-related deviations (call dropout, platform failure, rescheduling) were logged (Section 3.2), and no privacy or data-security incident occurred in either digital arm — directly relevant to the two principal safety questions for telerehabilitation, namely the absence of immediate physical assistance from the provider and the transmission of personal health information across digital platforms (Yau, et al., 2024, p. 2). Infection control in VRT followed the headset decontamination standard operating procedure (disposable face cover, documented post-use disinfection); no skin-contact or device-hygiene event was logged.

Within the threshold-based framework of this trial, tolerability was therefore complete in all three arms. The prevalence-style tolerability reporting recommended for VR clinical trials — covering physical symptoms (e.g., vertigo, nausea), emotional

symptoms (e.g., fear, anxiety), and equipment-related discomfort (e.g., ill-fitting headset, facial or nasal pain) (Birckhead, et al., 2018, p. 5) — was not part of the pre-specified outcome set, so sub-threshold experience cannot be expressed as prevalence; no discomfort event reached the deviation log in any case (Section 3.2). Adverse events per se, classified on severity, relatedness, and expectancy axes per ICH-GCP definitions (Section 2.2.3), are reported in Section 3.4.

3.4 Adverse Events

Adverse events were captured continuously across all 1,980 sessions (66 participants × 30 sessions) by direct therapist observation, app-based diary, weekly check-ins, and session metadata, and were classified per ICH-GCP definitions on severity, relatedness, and expectancy axes (Moussa, et al., 2025; Dennett, et al., 2022, p. 2), with serious events defined as death, life-threatening illness, hospitalization, persistent disability, or congenital anomaly (Moussa, et al., 2025, p. 2). Causality was adjudicated by an independent, allocation-blinded clinician on a five-point scale (Koh, et al., 2015, p. 12), and reporting followed the CONSORT Harms 2022 extension (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials, 2023).

No adverse events of any category occurred in any arm. Specifically, there were no serious adverse events; no falls during exercise (the pre-specified event of special interest); no exercise-attributable musculoskeletal flares; no exercise-stopping events — the individually reported category capturing any event leading to early session termination before planned protocol completion (Costa, et al., 2025, p. 19); no technology-related events (connectivity failure, audiovisual disturbance, privacy or data-security breach) in either digital arm; no threshold-level cybersickness in VRT (Section 3.3); and no headset-discomfort or device-hygiene events. With 22 participants per arm, the pre-specified threshold for individual listing — adverse events occurring in ≥5% of participants in either arm (Costa, et al., 2025, p. 19) — would have been crossed by any event affecting two or more participants in an arm; the observed zero therefore exhausts the CONSORT Harms reporting categories rather than omitting them (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials, 2023). The time-critical pathways to the sponsor and Data Safety Monitoring Board (within 24 hours) and the ethics committee (within 7 days) (Shaheed, Maher, Furmage, Hoffmann, & McLachlan, 2022; Wood, et al., 2024) were never triggered.

This zero-event result is consistent with the measured safety profile of both digital modes in the wider literature. Where telerehabilitation adverse events have been systematically captured, reported rates are low — 201 events across 65,352 sessions (0.31%; 3.1 per 1,000 sessions), predominantly physical, non-serious, and not directly attributed to the intervention (Shnitzer, et al., 2024) — and in home-based VR trials adverse events are predominantly mild and self-limited, with cybersickness the most commonly reported (Spiegel, et al., 2026). The present exposure is small relative to that benchmark: at the telerehabilitation rate, roughly six events would be expected across 1,980 sessions, so the observed zero documents a rate below that benchmark in this sample rather than establishing population-level safety. The explicit reporting of an absent event profile also addresses a known gap in rehabilitation research: a 2021 review of 249 trials of exercise for chronic low back pain found that only 12 measured adverse events in any systematic way (Shaheed, Maher, Furmage, Hoffmann, & McLachlan, 2022), the pattern CONSORT Harms 2022 was designed to correct (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials, 2023).

Two bounds qualify these findings. First, the safety conclusions were generated in a cohort partly selected for not requiring pharmacological pain management (analgesic use was an exclusion criterion; Section 2.1.1), so transportability to analgesic-using patients is undemonstrated (Section 4.1). Second, surveillance covered the 6-week intervention only — no post-intervention follow-up was scheduled (Section 2.2.3), so events arising after week 6 were not captured, the same constraint that bounds the durability conclusions (Section 4.1). Within these limits, the adverse-event record is complete and uniform across delivery modes: with zero events in every reporting category, no arm-specific safety table is added, and the deviation-level counts in Table 3 remain the definitive delivery-integrity record, closing the safety picture established in Sections 3.2 and 3.3.

4. Discussion

This trial tested whether three delivery modes of an identical, pre-specified exercise protocol — virtual reality therapy, telerehabilitation, and conventional physiotherapy — differ in 6-week disability reduction for adults with chronic low back pain. All arms improved substantially (mean ODI% changes of 24.91, 23.59, and 26.36 points; within-group $d > 5.0$; 100% attainment of the ≥10-point MCID), the primary ANCOVA did not detect a between-arm difference ($F = 2.709$, $p = 0.074$), and no adverse events of any category occurred. The pattern of uniform large gains with an undetected between-mode effect is consistent with evidence that remote rehabilitation can match face-to-face outcomes when content and dose are held constant (Comparison of remote rehabilitation and traditional

face-to-face rehabilitation for chronic low back pain: A systematic review and meta-analysis of randomized controlled trial, p. 2), with meta-analytic findings that VR confers no significant disability benefit over conventional care in CLBP (Opara, Kozinc, & Manojlović, 2024), and with the 100% adherence and retention observed here supporting good engagement with digitally mediated delivery (Garofano, et al., 2025; Shi, et al., 2024). Adverse events were captured deliberately and none occurred; where telerehabilitation events are measured, reported rates are low (201 events across 65,352 sessions; 0.31%; 3.1 per 1,000 sessions) (Shnitzer, et al., 2024).

4.1 Study Limitations

Nine limitations bound these findings.

1. Power and inferential framework. The trial was powered for a large omnibus effect only (Cohen's f = 0.40, $\alpha = 0.05$, $1-\beta = 0.80$; N = 66; achieved power 0.818), not for pairwise contrasts or for equivalence, and the observed omnibus effect ($\eta^2 = 0.058$) fell below this assumption. Margin-based designs in comparable CLBP trials have required 74 participants for non-inferiority, 102–104 for equivalence, and 120 enrolled allowing for dropout (Cui, et al., 2023; Zheng, et al., 2023, p. 7) — every figure exceeding the present N = 66. The non-significant result is therefore inconclusive rather than negative (Smith, et al., 2020; Manchikanti, 2008): a trial of this size cannot reliably exclude clinically important differences, and failure to find a difference is not proof of similarity (Liu, et al., 2010, p. 6).

2. No pre-specified margin. No non-inferiority or equivalence margin was pre-specified, so equivalence cannot be claimed; margin-based conclusions require a prospectively justified margin, one-sided testing, and power calculated against that margin (Flight & Julious, 2015, p. 2), with sample sizes roughly four times those of comparable difference-detecting trials (Jones, Jarvis, Lewis, & Ebbutt, 1996, p. 2) and the confidence interval evaluated against the pre-specified interval (Jones, Jarvis, Lewis, & Ebbutt, 1996; Piaggio, et al., 2012, p. 1). The exploratory post hoc TOST anchored to the ±10-point MCID is hypothesis-generating only, because the margin was defined after the data were observed and the sample was not sized for it (Linde & Ravenzwaaij, 2023; Campbell & Gustafson, 2021; Lauzon & Caffo, 2009, p. 4).

3. Sequential enrollment windows. The arms were enrolled in non-overlapping, sequential windows, perfectly confounding delivery mode with calendar time (seasonal variation, secular trends, and therapist experience maturation); no window-inclusive analysis was pre-specified or feasible, and screening-failure counts by exclusion criterion were not recorded, so differential selection across windows cannot be quantified. These between-arm findings must therefore be read as potentially confounded by calendar-time effects.

4. Selective cohort. Eligibility was restricted to adults aged 25–45 years who were medication-free (analgesic use was an exclusion criterion) and proficient on self-report digital-literacy measures, limiting generalizability to older (Oh, et al., 2021; Stamm, Dahms, & Müller-Werdan, 2020, p. 2) and less digitally experienced patients; the eHEALS is a measure validated in a youth web-based context rather than a test of headset-specific proficiency (Norman & Skinner, 2006, p. 1). The medication restriction is material to external validity: population-based studies report opioid use in 1.6%–18.8% of community CLBP samples, rising to 32.7% of patients in general practice and 64.4% in pain centres (Sola, Dueñas, Salazar, Ortega-Jiménez, & Failde, 2020, p. 14); Nordic registry cohorts document opioid use in 77.7% of CLBP patients and long-term NSAID use in 37.2% (Hallberg, et al., 2022, p. 8); and prevalence rises with episode recurrence, reaching 31.6%–48.0% for NSAIDs and 25.5%–44.2% for opioids among patients with 3–5 observed episodes (Ritzwoller, Crounse, Shetterly, & Rublee, 2006, p. 4).

5. Single 6-week endpoint. Outcomes were measured at a single endpoint with no post-intervention follow-up, so the durability of improvements in a recurrent condition is unknown: after resolution of an episode, roughly one-third of patients experience a recurrence within a year (Machado, et al., 2017), with previous episodes the only factor consistently predicting recurrence (Silva, et al., 2017). Meta-analytic evidence shows VR-based training advantages over conventional treatment were no longer statistically significant at short-term follow-up for disability (MD −1.28; 95% CI −4.47 to 1.90) (Li, et al., 2024, p. 10), and early home-based VR work found short-term efficacy without evidence of persistent benefit (Garrett, Taverner, & McDade, 2017) — while 6-month (García, et al., 2022, p. 1) and 12-month (Maddox, et al., 2025, p. 4) maintenance is detectable when observation windows extend. The two-timepoint design also precludes trajectory-level characterization of recovery.

6. No validated process measures. The Working Alliance Inventory — the most widely used alliance measure in musculoskeletal rehabilitation (Babatunde, MacDermid, & MacIntyre, 2017, p. 1) — and any presence/engagement scale were not administered in any arm, so mode-dependent differences in the therapeutic alliance, a predictor of CLBP outcomes (Ferreira, et al., 2012), cannot be quantified; higher alliance has been associated with substantial clinical benefit in telerehabilitation (McLaughlin, et al., 2023, p. 3), and presence is linked to VR-based analgesic efficacy (Triberti, Repetto, & Riva, 2014; Farič, et al., 2019; Micheluzzi, Vellone, & Iovino, 2024). Treatment

fidelity was high (30/30 sessions (Hildebrand, et al., 2012, p. 1)), but adherence is not relational equivalence: remote visual technique assessment was never validated against in-person examination, no inter-rater reliability assessment was conducted, and structured telerehabilitation assessments achieve only moderate inter-rater agreement (Richardson, Truter, Blumke, & Russell, 2016). Because a large extent of exercise-therapy improvement is attributable to contextual effects (Roode, Heymans, Lankveld, & Staal, 2024) and expectancy and credibility predict CLBP outcomes (Smeets, et al., 2008), attribution of the observed improvement to the exercise protocol rather than to delivery context is tempered (Comparison of remote rehabilitation and traditional face-to-face rehabilitation for chronic low back pain: A systematic review and meta-analysis of randomized controlled trial; Baroni, et al., 2023, p. 2).

7. Asymmetric screening and attention across arms. Only VRT underwent additional pre-randomization exclusion screening and the most intensive monitoring (cybersickness, falls risk, infection control) and therefore received the greatest volume of structured human and procedural contact; greater staff contact is associated with larger non-specific improvement (Smith, et al., 2020), and contextual effects (Roode, Heymans, Lankveld, & Staal, 2024) and expectancy (Smeets, et al., 2008) predict outcome — so the exploratory VRT-versus-TRT contrast may reflect differential screening or attention intensity rather than the immersive modality per se. In the present dataset the Tier C conversion pathway — which would have introduced treatment-received mixture into a digital arm — was never activated, so conversion-induced heterogeneity cannot have biased the estimates, although per-tier conversion counts and per-protocol sensitivity analyses remain reporting obligations for future trials (Schulz, Altman, Moher, & Group*, 2010, p. 1).

8. MCID ceiling and remaining bounds. Universal MCID attainment (22/22 in each arm; degenerate $\chi^2 = 0$, $p = 1.0$) rendered MCID non-discriminative — the smallest individual reduction (16.0 points) exceeded the ≥10-point threshold by 60%, and week-6 means (8.3–10.7 points) **fell within** the ≤20% "no-disability" region (Schwind, Learman, O'Halloran, Showalter, & Cook, 2013, p. 2). Headset-exposure minutes were not logged separately from session completion, so immersion-dose equivalence under the 20-minute cap cannot be verified post hoc, and no user-perception measures (e.g., Igroup Presence Questionnaire) were collected. Finally, single-country, four-site recruitment within the Faridabad–Gurugram National Capital Region constrains external validity to comparable Indian ambulatory-care settings.

9. On-site, therapist-attended delivery. All sessions in every arm — including the two digital arms — were delivered at the study facility with a therapist present throughout (same room for CPT; adjacent monitoring room for TRT/VRT; Section 2.1.4). The 100% adherence, 100% retention, and zero adverse events across all 30 sessions per participant were therefore secured by facility attendance and visible, accountable participation rather than by home-based engagement; unsupervised home-based telerehabilitation and VR would not carry this attendance incentive, and reported completion rates in sensor-based remote programs are 73%–90% even with gamification or real-time feedback (Garofano, et al., 2025). The feasibility results should not be extrapolated to unsupervised home use.

Strengths include assessor- and analyst-blinding, 100% retention and adherence across all three arms, an intervention identical in content and dose across modes, adverse events captured and reported explicitly per CONSORT Harms 2022 (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials, 2023), and reporting aligned with CONSORT 2010 (Schulz, Altman, Moher, & Group*, 2010, p. 1) and the TIDieR-telehealth extension (Rhon, et al., 2022).

5. Conclusion

In adults with chronic low back pain, an identical 6-week exercise protocol delivered as virtual reality therapy, telerehabilitation, or conventional physiotherapy produced large, clinically meaningful disability reductions at week 6: mean ODI% reductions of 24.91, 23.59, and 26.36 points, with within-group effect sizes $d > 5.0$ and 100% of participants in each arm attaining the pre-specified ≥10-point MCID (Vianin, 2008, p. 2). The pre-specified baseline-adjusted ANCOVA detected no statistically significant between-mode difference ($F = 2.709$, $p = 0.074$), with concordant sensitivity analyses (unadjusted ANOVA $F = 1.939$, $p = 0.152$, $\eta^2 = 0.058$), and no adverse events of any category occurred in any arm.

Because the trial was powered for difference detection (Cohen's $f = 0.40$; $N = 66$; achieved power 0.818) rather than for a pre-specified margin, this result constitutes insufficient evidence of a difference at six weeks — not evidence of equivalence, which would require a prospectively justified clinical margin, one-sided testing, and a margin-based sample size (Piaggio, et al., 2012; Flight & Julious, 2015). The exploratory TRT-versus-VRT contrast (−2.78 points; 95% CI −5.68 to 0.12; $d = -0.58$; raw $p = 0.060$) was not pre-specified, did not approach the Bonferroni-adjusted threshold ($\alpha = 0.0167$), and — consistent with meta-analytic evidence finding no significant pooled VR benefit for disability in CLBP (Opara, Kozinc, & Manojlović, 2024) — is best read as an imprecise numerical difference rather than evidence of a VRT advantage.

The pattern of uniformly large within-group improvement against a non-significant between-group comparison is consistent with the shared exercise protocol as the principal contributor to recovery and with the wider literature showing that digitally mediated delivery can match in-person outcomes when content and dose are held constant (Cui, et al., 2023, p. 1). From a translational perspective, these findings support the practical feasibility of TRT and VRT delivery in Indian ambulatory-care rehabilitation settings, particularly where physiotherapist-to-population ratios and travel burden limit access; however, VRT and TRT should be described as candidates for further evaluation rather than established substitutes for conventional physiotherapy, and adoption should await adequately powered, margin-based confirmation. Future trials should pre-specify non-inferiority or equivalence margins anchored to the clinically meaningful ODI difference, extend follow-up beyond six weeks to establish durability in a recurrent condition, broaden eligibility to older and medication-using populations, and apply more stringent responder thresholds to enhance discriminative sensitivity across delivery modes.

Declarations

Trial registration: CTRI/2023/11/059788 (Clinical Trials Registry – India), registered prospectively on 14 November 2023. Ethics approval: GDGU Research Ethics Committee (GDGU/SoMAS/2023/168A) and Waves Women Empowerment Trust Independent Ethics Committee (WWET/2023/IEC-AP/04); written informed consent was obtained from all participants. Funding: self-funded by the principal investigator; no external funding received. Conflicts of interest: none declared. Data availability: de-identified participant data available from the corresponding author upon reasonable request. Conflicts of interest: The corresponding author is Bharat Bhushan of GD Goenka University, Gurugram, which served as one of the four recruitment sites. The author had no role in allocation, outcome assessment, or statistical analysis. All other authors declare no competing interests. Reporting guidelines: CONSORT 2010 (Group, Schulz, Altman, & Moher, 2010), CONSORT Harms 2022 (Junqueira, et al., CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomized trials, 2023), TIDieR (Hoffmann, et al., 2014), and TIDieR-Telehealth (Rhon, et al., 2022), with the non-inferiority/equivalence extension applied as an interpretive standard (Piaggio, et al., 2012; Hopewell, et al., 2016).

Table 4. Intervention Reporting per TIDieR Checklist and the TIDieR-Telehealth Extension

TIDieR Item	CPT (In-Person)	TRT (Teams Videoconference)	VRT (Utopia 360 Headset)
1. Brief name	In-person, therapist-supervised exercise	Synchronous videoconference-delivered exercise via Microsoft Teams	Therapist-guided exercise via Utopia 360 VR headset with remote supervision
2. Rationale	Evidence-based standard of care; direct physical correction and hands-on cueing	To expand access, reduce travel burden, and validate remote delivery of an effective in-person intervention	To combine remote-access benefits with immersive, first-person visual/motor learning
3. Materials	Clinic space, mats, standard equipment, printed handouts	Patient-owned smartphone, tablet, or laptop; Microsoft Teams application; stable internet connection; joining-link instructions	Utopia 360 VR headset, compatible smartphone, VR exercise app, remote-monitoring link
4. Procedures	Therapist demonstrates and physically corrects technique in real time	The therapist observes via the Teams feed, provides verbal cues, and corrects form via webcam	The therapist connects remotely to observe and guide the patient; the patient views a 360° exercise demonstration in the headset; the therapist provides corrections through audio cueing
5. Who provided	Registered physiotherapist, standard in-person training	Registered physiotherapist, additional videoconferencing/remote-assessment competency	Registered physiotherapist, additional VR-platform/remote-monitoring competency

6. How (mode)	Face-to-face, individual, synchronous	Remote, individual, synchronous, bidirectional (audio and video)	Remote, individual, synchronous; audio (± video) cueing while immersed
7. Where	Clinic treatment room (single location)	Two locations — patient room + therapist remote site, via Teams	Two locations — patient room (space/safety-checked) + therapist remote monitoring site
8. Timing and dose	5 sessions/week × 6 weeks, in-person	5 sessions/week × 6 weeks, via Teams	5 sessions/week × 6 weeks, via VR headset
9. Tailoring	Progressed via direct in-person assessment	Tailored via self-report + webcam visual assessment (no palpation)	Tailored via self-report + remote monitoring; VR content/difficulty adjusted by therapist
10. Modifications	In-session changes documented (what/why/when/how)	Platform/session-format/connectivity workarounds documented	VR content/headset-setup/connectivity workarounds documented
11. Fidelity (planned)	In-person checklist/observation by senior physiotherapist	Session-recording review/clinician self-report checklist	Clinician self-report checklist + verification of adequate remote observation
12. Fidelity (actual)	30/30 sessions completed as planned (100%); 0 missed sessions	30/30 sessions completed as planned (100%); 0 technology deviations logged	30/30 sessions completed as planned (100%); 0 headset/connectivity/cybersickness deviations logged

References

1. Wan F. Statistical analysis of two arm randomized pre-post designs with one post-treatment measurement. 2021 July 24 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s12874-021-01323-9
2. Piaggio G, Elbourne DR, Pocock S, Evans S, Altman DG, Group for the C. Reporting of Noninferiority and Equivalence Randomized Trials. 2012 Dec 26 [cited 2026 Jan]; Available from: https://doi.org/10.1001/jama.2012.87802
3. Flight L, Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. 2015 Nov 25 [cited 2025 Nov]; Available from: https://doi.org/10.1002/pst.1716
4. Opara M, Kozinc Ž, Manojlović D. Effects of virtual reality on pain, disability and kinesiophobia in patients with chronic low back pain: A systematic review with meta-analysis. 2024 Aug 17 [cited 2025 Nov]; Available from: https://doi.org/10.3233/ppr-230835
5. Tagliaferri SD, Miller CT, Owen PJ, Mitchell UH, Brisby H, Fitzgibbon BM, et al. Domains of Chronic Low Back Pain and Assessing Treatment Effectiveness: A Clinical Perspective. 2019 Oct 14 [cited 2026 Mar]; Available from: https://doi.org/10.1111/papr.12846
6. Slatman S, Ostelo R, Goor H van, Staal JB, Knoop J. Physiotherapy with integrated virtual reality for patients with complex chronic low back pain: protocol for a pragmatic cluster randomized controlled trial (VARIETY study). 2023 Feb 20 [cited 2026 Mar]; Available from: https://doi.org/10.1186/s12891-023-06232-0
7. Cui D, Janela D, Costa F, Molinos M, Areias AC, Moulder R, et al. Randomized-controlled trial assessing a digital care program versus conventional physiotherapy for chronic low back pain. 2023 July 7 [cited 2026 Mar]; Available from: https://doi.org/10.1038/s41746-023-00870-3
8. Garofano M, Sorbo RD, Calabrese M, Giordano M, Palo MPD, Bartolomeo MD, et al. Remote Rehabilitation and Virtual Reality Interventions Using Motion Sensors for Chronic Low Back Pain: A Systematic Review of Biomechanical, Pain, Quality of Life, and Adherence Outcomes. 2025 May 6 [cited 2025 Nov]; Available from: https://doi.org/10.3390/technologies13050186
9. The Effects of Virtual Immersive Gaming to Optimize Recovery (VIGOR) in Low Back Pain: A Phase II Randomized Controlled Trial [Internet]. [cited 2026 Mar]. Available from: https://doi.org/10.3390/healthcare14020142
10. Shi W, Zhang Y, Bian Y, Chen L, Yuan W, Zhang H, et al. The Physical and

Psychological Effects of Telerehabilitation-Based Exercise for Patients With Nonspecific Low Back Pain: Prospective Randomized Controlled Trial. 2024 July 9 [cited 2026 Jan]; Available from: https://doi.org/10.2196/56580

11. Zheng F, Li Z, Huanjie H, Zhang X, Chen R, Jiahui P, et al. Home-based rehabilitation training with human key point detection for chronic low back pain patients: a randomized controlled trial protocol. 2023 Nov 27 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s13063-023-07805-z
12. Comparison of remote rehabilitation and traditional face-to-face rehabilitation for chronic low back pain: A systematic review and meta-analysis of randomized controlled trial [Internet]. [cited 2026 June]. Available from: https://doi.org/10.36922/ejmo026090102
13. Lawford BJ, Bennell KL, Campbell PK, Kasza J, Hinman RS. Association Between Therapeutic Alliance and Outcomes Following Telephone-Delivered Exercise by a Physical Therapist for People With Knee Osteoarthritis: Secondary Analyses From a Randomized Controlled Trial. 2020 Dec 19 [cited 2026 Jan]; Available from: https://doi.org/10.2196/23386
14. Ferreira PH, Ferreira ML, Maher CG, Refshauge KM, Latimer J, Adams R. The Therapeutic Alliance Between Clinicians and Patients Predicts Outcome in Chronic Low Back Pain. 2012 Nov 9 [cited 2025 Nov]; Available from: https://doi.org/10.2522/ptj.20120137
15. Prel J du, Hommel G, Röhrig B, Blettner M. Confidence Interval or P-Value? Part 4 of a Series on Evaluation of Scientific Publications. 2009 May 8 [cited 2026 Jan]; Available from: https://doi.org/10.3238/arztebl.2009.0335
16. Vianin M. Psychometric properties and clinical usefulness of the Oswestry Disability Index. 2008 Nov 13 [cited 2026 Jan]; Available from: https://doi.org/10.1016/j.jcm.2008.07.001
17. Schulz KF, Altman DG, Moher D, Group* for the C. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. 2010 June 1 [cited 2025 Nov]; Available from: https://doi.org/10.7326/0003-4819-152-11-201006010-00232
18. Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. 2014 Mar 7 [cited 2026 Sept]; Available from: https://doi.org/10.1136/bmj.g1687
19. Rhon DI, Fritz JM, Kerns RD, McGeary DD, Coleman BC, Farrokhi S, et al. TIDieR-telehealth: precision in reporting of telehealth interventions used in clinical trials - unique considerations for the Template for the Intervention Description and Replication (TIDieR) checklist. 2022 June 2 [cited 2026 Aug]; Available from: https://doi.org/10.1186/s12874-022-01640-7
20. Cashin AG, Lee H, Bagg MK, O'Hagan E, Traeger AC, Kamper SJ, et al. A systematic review highlights the need to improve the quality and applicability of trials of physical therapy interventions for low back pain. 2020 June 27 [cited 2026 Jan]; Available from: https://doi.org/10.1016/j.jclinepi.2020.06.025
21. Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. 2019 Jan 11 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s13063-018-3108-3
22. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. 2014 Apr 9 [cited 2026 Jan]; Available from: https://doi.org/10.1186/1471-2288-14-49
23. Shatz I. Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. 2023 Mar 3 [cited 2026 Jan]; Available from: https://doi.org/10.3758/s13428-023-02072-x
24. Zygmont CS. Managing the Assumption of Normality within the General Linear Model with Small Samples: Guidelines for Researchers Regarding If, When and How. 2023 Dec 1 [cited 2026 Jan]; Available from: https://doi.org/10.20982/tqmp.19.4.p302
25. Baroni MP, Jacob MFA, Rios WRC, Fandim JV, Fernandes LG, Chaves PI, et al. The state of the art in telerehabilitation for musculoskeletal conditions. 2023 Jan 4 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s40945-022-00155-0
26. Hildebrand M, Host HH, Binder EF, Carpenter B, Freedland KE, Morrow-Howell N, et al. Measuring Treatment

Fidelity in a Rehabilitation Intervention Study. 2012 Feb 29 [cited 2026 Jan]; Available from: https://doi.org/10.1097/phm.0b013e31824ad462
27. Dennett AM, Harding KE, Peiris CL, Shields N, Barton CJ, Lynch L, et al. Efficacy of Group Exercise–Based Cancer Rehabilitation Delivered via Telehealth (TeleCaRe): Protocol for a Randomized Controlled Trial. 2022 July 18 [cited 2026 Jan]; Available from: https://doi.org/10.2196/38553
28. Koh GC, Yen SC, Tay A, Cheong A, Ng YS, Silva DAD, et al. Singapore Tele-technology Aided Rehabilitation in Stroke (STARS) trial: protocol of a randomized clinical trial on tele-rehabilitation for stroke patients. 2015 Sept 4 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s12883-015-0420-3
29. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. 2006 Nov 14 [cited 2026 Sept]; Available from: https://doi.org/10.2196/jmir.8.4.e27
30. Moussa R, Power J, Yenson V, Fazekas B, Marston C, Hosie A, et al. Improving the capture and reporting of adverse events in clinical trials of non-pharmacological interventions: learnings from the PaCCSC/CST membership. 2025 Mar 18 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s13063-025-08801-1
31. Costa A dos S, Barbosa CB, Guizilini S, Begot I, Moraes PI de M, Santos VR dos, et al. REVIVE-HF: Rehabilitation with Immersive Virtual Reality and Exercise in Hospitalized Patients with Heart Failure - A Randomized Controlled Trial Protocol. 2025 Dec 15 [cited 2026 May]; Available from: https://doi.org/10.64898/2025.12.11.25342100
32. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. 2018 Oct 23 [cited 2026 Mar]; Available from: https://doi.org/10.2196/11973
33. Yau T, McIntyre M, Chan J, Bhogal D, Andreoli A, Bayley M, et al. Adverse events associated with the delivery of telerehabilitation: A scoping review protocol. 2024 Feb 21 [cited 2026 Jan]; Available from: https://doi.org/10.1371/journal.pone.0297908
34. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, Usability, and Enjoyment of a Home-Based Exercise Program Delivered via an Exercise App for Musculoskeletal Health in Community-Dwelling Older Adults: Short-term Prospective Pilot Study. 2020 Oct 29 [cited 2026 Jan]; Available from: https://doi.org/10.2196/21094
35. Junqueira DR, Zorzela L, Golder S, Loke YK, Gagnier J, Julious SA, et al. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. 2023 Apr 24 [cited 2025 Dec]; Available from: https://doi.org/10.1136/bmj-2022-073725
36. Shaheed CA, Maher CG, Furmage A, Hoffmann T, McLachlan AJ. Strengthening the reporting of harms of all interventions in clinical trials. 2022 Oct 26 [cited 2025 Nov]; Available from: https://doi.org/10.5694/mja2.51755
37. Wood G, Scheer A, Saundankar J, Tran D, Cordina R, Maiorana A. The effects of telerehabilitation in adults with complex biventricular congenital heart conditions: protocol for a multi-centre, randomised controlled trial—CH-FIT. 2024 Apr 5 [cited 2026 Feb]; Available from: https://doi.org/10.1186/s13063-024-08019-7
38. Shnitzer H, Chan J, Yau T, McIntyre M, Andreoli A, Kua A, et al. The Safety of Telerehabilitation: A Systematic Review (Preprint). 2024 Nov 14 [cited 2025 Dec]; Available from: https://doi.org/10.2196/preprints.68681
39. Alrwaily M. Delivering Musculoskeletal Rehabilitation in the Digital Era: A Perspective on Clinical Strategies for Remote Practice. 2025 Sept 12 [cited 2025 Dec]; Available from: https://doi.org/10.3390/healthcare13182286
40. Russell T, Truter P, Blumke R, Richardson B. The Diagnostic Accuracy of Telerehabilitation for Nonarticular Lower-Limb Musculoskeletal Disorders. 2010 June 1 [cited 2025 Dec]; Available from: https://doi.org/10.1089/tmj.2009.0163
41. Keshavarz B, Hecht H. Validating an Efficient Method to Quantify Motion Sickness. 2011 Apr 11 [cited 2026 Aug]; Available from: https://doi.org/10.1177/0018720811403736

42. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. 1993 July 1 [cited 2026 Sept]; Available from: https://doi.org/10.1207/s15327108ijap0303_3
43. Lusardi MM, Fritz SL, Middleton A, Allison LK, Wingood M, Phillips E, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. 2016 Aug 18 [cited 2025 Nov]; Available from: https://doi.org/10.1519/jpt.0000000000000099
44. Palumbo P, Palmerini L, Bandinelli S, Chiari L. Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better? 2015 Dec 30 [cited 2026 Jan]; Available from: https://doi.org/10.1371/journal.pone.0146247
45. Pillay J, Gaudet L, Saba S, Vandermeer B, Ashiq AR, Wingert A, et al. Falls prevention interventions for community-dwelling older adults: systematic review and meta-analysis of benefits, harms, and patient values and preferences. 2024 Nov 26 [cited 2026 Mar]; Available from: https://doi.org/10.1186/s13643-024-02681-3
46. Bischoff H. Identifying a cut-off point for normal mobility: a comparison of the timed "up and go" test in community-dwelling and institutionalised elderly women. 2003 Apr 29 [cited 2026 Jan]; Available from: https://doi.org/10.1093/ageing/32.3.315
47. Schoene D, Wu SM -S., Mikolaizak AS, Menant JC, Smith S, Delbaere K, et al. Discriminative Ability and Predictive Validity of the Timed Up and Go Test in Identifying Older People Who Fall: Systematic Review and Meta-Analysis. 2013 Jan 25 [cited 2025 Nov]; Available from: https://doi.org/10.1111/jgs.12106
48. Visual feedback and guided balance training in an immersive virtual reality environment for lower extremity rehabilitation [Internet]. [cited 2026 May]. Available from: https://doi.org/10.1016/j.cag.2024.01.007
49. Stamm O, Dahms R, Müller-Werdan U. Virtual reality in pain therapy: a requirements analysis for older adults with chronic back pain. 2020 Sept 29 [cited 2026 Mar]; Available from: https://doi.org/10.1186/s12984-020-00753-8
50. Lauzon CB, Caffo B. Easy Multiplicity Control in Equivalence Testing Using Two One-Sided Tests. 2009 May 1 [cited 2026 Jan]; Available from: https://doi.org/10.1198/tast.2009.0029
51. Linde M, Ravenzwaaij D van. baymedr: an R package and web application for the calculation of Bayes factors for superiority, equivalence, and non-inferiority designs. 2023 Nov 24 [cited 2026 Mar]; Available from: https://doi.org/10.1186/s12874-023-02097-y
52. Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. 2013 May 1 [cited 2026 Jan]; Available from: https://doi.org/10.1179/2042618613y.0000000028
53. Smith SM, Fava M, Jensen MP, Mbowe O, McDermott MP, Turk DC, et al. John D. Loeser Award Lecture: Size does matter, but it isn't everything: the challenge of modest treatment effects in chronic pain clinical trials. 2020 Aug 15 [cited 2025 Nov]; Available from: https://doi.org/10.1097/j.pain.0000000000001849
54. Manchikanti L. Evidence-Based Medicine, SystematicReviews, and Guidelines in Interventional PainManagement: Part 2: Randomized ControlledTrials. 2008 Dec 14 [cited 2025 Nov]; Available from: https://doi.org/10.36076/ppj.2008/11/717
55. Spiegel BM, Eberlein S, Persky S, Ishimori M, Venuturupalli S, Tu J, et al. Randomized-controlled trial of skills-based vr vs. distraction vr vs. sham VR for chronic low back pain. 2026 Feb 16 [cited 2026 Mar]; Available from: https://doi.org/10.1038/s41746-026-02437-4
56. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley WT, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. 2010 Aug 6 [cited 2026 Jan]; Available from: https://doi.org/10.1016/j.jclinepi.2009.11.021
57. Jones B, Jarvis P, Lewis JA, Ebbutt A. Trials to assess equivalence: the importance of rigorous methods: Fig 1. 1996 July 6 [cited 2026 Jan]; Available from: https://doi.org/10.1136/bmj.313.7048.36
58. Campbell H, Gustafson P. What to make of equivalence testing with a post-specified

margin? 2021 Sept 6 [cited 2025 Nov]; Available from: https://doi.org/10.15626/mp.2020.2506
59. Oh SS, Kim K, Kim M, Oh J, Chu SH, Choi J. Measurement of Digital Literacy Among Older Adults: Systematic Review. 2021 Jan 16 [cited 2026 Jan]; Available from: https://doi.org/10.2196/26145
60. Sola HD, Dueñas M, Salazar A, Ortega-Jiménez P, Failde I. Prevalence of Therapeutic use of Opioids in Chronic non-Cancer Pain Patients and Associated Factors: A Systematic Review and Meta-Analysis. 2020 Nov 18 [cited 2026 Jan]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7750787/pdf
61. Hallberg S, Rolfson O, Karppinen J, Schiøttz-Christensen B, Stubhaug A, Fischer MR, et al. Burden of disease and management of osteoarthritis and chronic low back pain: healthcare utilization and sick leave in Sweden, Norway, Finland and Denmark (BISCUITS): study design and patient characteristics of a real world data study. 2022 July 20 [cited 2025 Dec]; Available from: https://portal.findresearcher.sdu.dk/da/publications/1ca120fa-7d90-4b45-8186-b84406839d0c
62. Ritzwoller DP, Crounse L, Shetterly S, Rublee D. The association of comorbidities, utilization and costs for patients identified with low back pain. 2006 Sept 18 [cited 2026 Jan]; Available from: https://doi.org/10.1186/1471-2474-7-72
63. Machado GC, Maher CG, Ferreira PH, Latimer J, Koes BW, Steffens D, et al. Can Recurrence After an Acute Episode of Low Back Pain Be Predicted? 2017 June 15 [cited 2025 Nov]; Available from: https://doi.org/10.1093/ptj/pzx067
64. Silva T da, Mills K, Brown BT, Herbert R, Maher CG, Hancock MJ. Risk of Recurrence of Low Back Pain: A Systematic Review. 2017 Mar 30 [cited 2025 Nov]; Available from: https://doi.org/10.2519/jospt.2017.7415
65. Li R, Li Y, Kong Y, Li H, Hu D, Fu C, et al. Virtual Reality–Based Training in Chronic Low Back Pain: Systematic Review and Meta-Analysis of Randomized Controlled Trials. 2024 Feb 26 [cited 2026 Mar]; Available from: https://doi.org/10.2196/45406
66. Garrett B, Taverner T, McDade P. Virtual Reality as an Adjunct Home Therapy in Chronic Pain Management: An Exploratory Study. 2017 May 11 [cited 2026 Mar]; Available from: https://doi.org/10.2196/medinform.7271
67. García L, Birckhead B, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, et al. Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: 6-Month Follow-up Study of a Randomized Clinical Trial. 2022 May 25 [cited 2026 Mar]; Available from: https://doi.org/10.2196/37480
68. Maddox T, Oldstone L, Linde-Zwirble W, Bonakdar RA, Maddox R, Sackman J, et al. Differential treatment response to virtual reality in high-impact chronic pain: secondary analysis of a randomized trial. 2025 Apr 25 [cited 2026 June]; Available from: https://doi.org/10.1038/s41598-025-98716-3
69. Babatunde FO, MacDermid JC, MacIntyre NJ. Characteristics of therapeutic alliance in musculoskeletal physiotherapy and occupational therapy practice: a scoping review of the literature. 2017 May 30 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s12913-017-2311-3
70. McLaughlin KH, Fritz JM, Minick KI, Brennan GP, McGee T, Lane E, et al. Examining the Relationship Between Individual Patient Factors and Substantial Clinical Benefit From Telerehabilitation Among Patients With Chronic Low Back Pain. 2023 Dec 29 [cited 2026 June]; Available from: https://doi.org/10.1093/ptj/pzad180
71. Triberti S, Repetto C, Riva G. Psychological Factors Influencing the Effectiveness of Virtual Reality–Based Analgesia: A Systematic Review. 2014 June 1 [cited 2025 Dec]; Available from: https://doi.org/10.1089/cyber.2014.0054
72. Farič N, Potts H, Hon A, Smith L, Newby K, Steptoe A, et al. What Players of Virtual Reality Exercise Games Want: Thematic Analysis of Web-Based Reviews. 2019 Apr 18 [cited 2026 Mar]; Available from: https://doi.org/10.2196/13833
73. Micheluzzi V, Vellone E, Iovino P. A Situation-Specific Theory on the Use of Immersive Virtual Reality in Rehabilitation for Patients With Disabilities. 2024 Dec 12 [cited 2026 June]; Available from: https://doi.org/10.1097/hnp.0000000000000718
74. Richardson BR, Truter P, Blumke R, Russell T. Physiotherapy assessment and diagnosis of musculoskeletal disorders of the knee via telerehabilitation. 2016 Mar 17 [cited 2025 Nov]; Available from:

https://doi.org/10.1177/1357633x15627237

75. Roode A de, Heymans MW, Lankveld W van, Staal JB. The impact of contextual effects in exercise therapy for low back pain: a systematic review and meta-analysis. 2024 Oct 23 [cited 2026 Mar]; Available from: https://doi.org/10.1186/s12916-024-03679-3
76. Smeets R, Beelen S, Goossens MEJB, Schouten EG, Knottnerus JA, Vlaeyen JWS. Treatment Expectancy and Credibility Are Associated With the Outcome of Both Physical and Cognitive-behavioral Treatment in Chronic Low Back Pain. 2008 May 1 [cited 2026 Jan]; Available from: https://doi.org/10.1097/ajp.0b013e318164aa75
77. Group the C, Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. 2010 Mar 24 [cited 2026 Sept]; Available from: https://doi.org/10.1186/1741-7015-8-18
78. Junqueira DR, Zorzela L, Golder S, Loke YK, Gagnier J, Julious SA, et al. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomized trials. 2023 Apr 24 [cited 2026 Aug]; Available from: https://doi.org/10.1016/j.jclinepi.2023.04.005
79. Hopewell S, Boutron I, Altman DG, Barbour V, Moher D, Montori VM, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. 2016 Nov 28 [cited 2026 Jan]; Available from: https://doi.org/10.1186/s12916-016-0736-x