Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Similar Mood Benefit, Unequal Retention: A Pooled Analysis of Two Pilot Randomized Trials of Virtual Reality Boxing in Autistic and Non-Autistic Young Adults

Youngjun Lee

leey14@chop.edu

Michigan State University

Research Article

Keywords: Autism, Virtual reality, Exercise, Depressive symptoms, Heart rate variability, Retention

Posted Date: October 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11394857/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Exercise reduces depressive symptoms in adults, but little is known about whether autistic adults respond to the same exercise program, or stay in it, to the same degree as non-autistic peers. This secondary analysis pooled individual participant data from two pilot randomized trials of virtual reality (VR) boxing that shared a laboratory, outcome measures, and a no-exercise control condition: one with non-autistic university students (*n* = 30; six weekly sessions) and one with autistic university students (*n* = 20; eight sessions over 4 weeks). Baseline-adjusted VR − control differences were estimated within each trial and compared through a group × population interaction. At baseline, autistic participants had lower resting heart rate variability (HRV), more daily sitting time, and more visceral fat. The reduction in depressive symptoms on the PHQ-9 was nearly identical in the two trials (non-autistic: −1.99 points, 95% CI [− 4.36, 0.39]; autistic: −2.19, 95% CI [− 6.57, 2.19]; difference − 0.21, 95% CI [− 4.68, 4.26]), and this held across all sensitivity analyses. Retention did not: all 16 non-autistic participants assigned to VR boxing completed post-testing, compared with 6 of 11 autistic participants (p = .006), and four of the five autistic non-completers stopped after the first session. In exploratory analyses, resting HRV and sitting time changed in opposite directions in the two trials. Autistic young adults who completed VR boxing showed mood benefits comparable to those of non-autistic peers, but nearly half did not continue past the first session, which makes onboarding a design priority for future trials.

Introduction

Depressive disorders are among the most frequent co-occurring diagnoses in autism (Lai et al., 2019), and roughly one in four autistic adults currently meets criteria for depression (Hollocks et al., 2019). In non-autistic adults, exercise is an effective treatment for depressive symptoms, with effects in the range of established psychological and pharmacological treatments (Noetel et al., 2024; Schuch et al., 2016). Whether the same holds for autistic adults is much less clear, partly because autistic people are less likely to take part in structured physical activity in the first place. Studies of autistic children and adults point to sensory sensitivities, the social demands of group settings, unpredictable routines, and motor coordination difficulties as barriers to exercise (Hillier et al., 2020; Must et al., 2015; Obrusnikova & Cavalier, 2011). Autistic adults also have higher rates of obesity, diabetes, and hypertension than non-autistic adults (Croen et al., 2015), which adds to the case for finding forms of exercise they will actually do.

Immersive virtual reality (VR) offers one way around some of these barriers. A head-mounted display lets a person exercise alone, at home or in a quiet room, with visual and auditory input that can be adjusted and routines that repeat predictably. Autistic users generally accept head-mounted displays, although sensory discomfort and cybersickness come up repeatedly (Bradley & Newbutt, 2018). Boxing workouts suit this format well. They consist of choreographed punch, block, and dodge sequences cued to music, they involve no physical contact, and they require no partner. Boxing-based training has also been studied as an exercise therapy in Parkinson's disease, mostly for balance, mobility, and quality of life (Wang et al., 2025).

A practical question for anyone designing exercise programs for autistic adults is whether they respond to a program the way non-autistic adults do. Single-population trials cannot answer it. If a trial in autistic adults finds a small effect, the result may reflect something about autism, about the protocol, or about the sample, and a separate trial in non-autistic adults run with a different protocol does not settle which. A direct comparison needs the same intervention, the same measures, and a randomized control group within each population. Two outcomes deserve attention. One is the effect among people who receive the intervention. The other is whether people stay long enough to receive it, which is the main question pilot trials are meant to answer (Eldridge et al., 2016; Leon et al., 2011).

Heart rate variability (HRV) adds a physiological angle. Resting HRV, particularly vagally mediated indices such as the root mean square of successive differences (RMSSD), is thought to index the capacity to regulate emotion (Thayer & Lane, 2000), and autistic individuals show lower resting HRV than non-autistic individuals on average (Cheng et al., 2020). In a pilot trial of VR boxing with non-autistic university students, depressive symptoms improved while resting HRV did not, a pattern the authors described as an affective–autonomic dissociation (Lee, Vazou, Kim, & Hauck, 2026). Whether autistic adults, who start from a lower autonomic baseline, show the same dissociation has not been tested.

This study pooled individual participant data from that trial and from a pilot trial of VR boxing in autistic university students run in the same laboratory with the same measures. The aims were to (a) compare the baseline mental health, autonomic, and body composition profiles of the two samples; (b) estimate whether the effect of VR boxing on depressive symptoms (primary outcome), resting HRV, sleep quality, and resting heart rate differed between autistic and non-autistic participants; and (c) compare retention and adherence. Given the sample sizes, the analysis was designed for estimation and hypothesis generation rather than confirmatory testing.

Methods

Design and Data Sources

Both trials were two-arm, parallel-group pilot randomized trials conducted in 2025 in the same university kinesiology laboratory in the Midwestern United States under a single institutional review board protocol. The non-autistic trial compared six weekly VR boxing sessions with a no-exercise control condition; its design and primary outcomes are reported in Lee, Vazou, Kim, and Hauck (2026). The autism trial compared eight VR boxing sessions over 4 weeks with a no-exercise control condition; its primary outcomes are reported separately (manuscript under review). Both trials used the same baseline and post-test battery. The pooled comparison reported here was planned after both trials had been completed and was not specified in either protocol. Neither trial was prospectively registered in a public trial registry, because both were designed as small pilot studies to estimate feasibility parameters and effect sizes for a future definitive trial. Reporting follows the CONSORT extension for pilot and feasibility trials where it applies (Eldridge et al., 2016).

Participants

The non-autistic trial enrolled 30 university students aged 18–24 years with no neurodevelopmental diagnosis, recruited through university email bulletins, flyers, and social media. Exclusion criteria were medical conditions for which exercise is commonly discouraged, pregnancy, and conditions likely to be worsened by VR use (Lee, Vazou, Kim, & Hauck, 2026).

Participants in the autism trial were recruited from undergraduates who completed an online screening survey about interest in a 4-week VR boxing program (Lee, Vazou, Kim, Ratan, et al., 2026). Of 179 valid respondents, 58 reported a formal autism diagnosis, and 20 of these enrolled and were randomized (11 to VR boxing, 9 to control). Eligibility required a self-reported autism diagnosis confirmed by a parent or caregiver, fluent English, the ability to perform light-to-moderate exercise, and no contraindication to VR use such as photosensitive epilepsy. Participants taking psychiatric medication had been on a stable regimen for at least 4 weeks before enrollment. Participant flow for both trials is shown in Fig. 1.

In both trials, participants were randomly allocated after baseline testing. In the non-autistic trial, allocation used a computer-generated sequence stratified by sex and was concealed from researchers until baseline testing was complete (Lee, Vazou, Kim, & Hauck, 2026). All participants provided informed consent; in the autism trial, consent was obtained electronically.

Interventions

In both trials, the VR boxing condition used commercially available boxing workouts on a standalone head-mounted display (in the autism trial, Supernatural on Meta Quest 3). Workouts consisted of punch, block, and dodge targets timed to music. Non-autistic participants completed one laboratory session per week for 6 weeks while wearing a heart rate band. Autistic participants completed two sessions per week for 4 weeks (eight sessions of about 34 min, each with a warm-up, a choreographed boxing routine, and a cool-down) in the laboratory or at home, according to their preference. A researcher supervised every session, including those at home, and participants chose the workout intensity and music genre from preset options; some attended laboratory sessions with a parent. Control participants in both trials kept their usual routines and were asked not to start any new structured exercise. Post-testing took place 6 weeks after baseline in the non-autistic trial and at the end of the 6-week study period (baseline week, 4-week intervention, post-test week) in the autism trial.

Measures

Depressive symptoms were measured with the nine-item Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), which covers the past 2 weeks and ranges from 0 to 27. Resting HRV was recorded at each visit after participants had sat quietly for 5 min; a 5-min recording was then taken with a Polar H10 chest strap (Polar Electro, Finland), and RMSSD was extracted with Elite HRV software and natural-log transformed (lnRMSSD; Laborde et al., 2017). Resting heart rate came from the same recording, and the software's signal-quality rating was kept for a sensitivity analysis. Sleep quality was measured with the

Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989; range 0–21). Physical activity and sitting time were measured with the Global Physical Activity Questionnaire (GPAQ; Bull et al., 2009). Activity was expressed in MET-minutes per week using the standard weights (8 for vigorous work and recreation; 4 for moderate work, moderate recreation, and active transport), and sitting time in minutes per day. Body composition was measured by bioelectrical impedance (InBody 570, InBody Co., Seoul, Korea), from which BMI, body fat percentage, skeletal muscle index, and visceral fat level were taken.

Feasibility outcomes were retention (completion of the post-test), adherence (sessions attended out of sessions offered), and session duration. Exercise intensity was recorded as mean session heart rate in the non-autistic trial and as the in-app difficulty setting in the autism trial. In the autism trial, the Simulator Sickness Questionnaire (Kennedy et al., 1993) was completed after the first session, which allowed identification of participants who stopped after that session.

Statistical Analysis

The quantity of interest was the difference between trials in the baseline-adjusted effect of VR boxing relative to control. Within each trial, this effect was estimated with an analysis of covariance of the post-test score on group and the baseline score. The two trials were then analyzed together in a one-stage individual participant data model with the post-test score as the outcome and baseline score, population, baseline × population, group, and group × population as predictors. The group × population coefficient is the difference in VR effects (autism trial minus non-autistic trial). Trial-specific intercepts and baseline slopes keep the comparison within trials, as recommended for treatment–covariate interactions in pooled trial data (Riley et al., 2020). Because population and trial are the same variable in these data, this coefficient compares two randomized effects across trials and is not itself a randomized contrast.

Ten outcomes were analyzed: depressive symptoms (primary); lnRMSSD, sleep quality, and resting heart rate (secondary); and body fat percentage, skeletal muscle index, visceral fat level, BMI, physical activity, and sitting time (exploratory). Holm's (1979) procedure was applied across the ten interaction tests. Standardized effects were obtained by dividing each adjusted difference by the baseline standard deviation of the respective trial. With samples of this size, only large interactions can be detected (Brookes et al., 2004), so the minimum detectable interaction at 80% power is reported alongside the estimates.

The primary analysis used complete cases, as in the original trial reports. Sensitivity analyses comprised (a) HC3 heteroskedasticity-consistent standard errors (Long & Ervin, 2000); (b) a two-stage analysis in which the trial-specific estimates were differenced, with Welch–Satterthwaite degrees of freedom; (c) an intention-to-treat analysis using multiple imputation by chained equations with predictive mean matching, run separately within each trial (200 imputations; predictors: baseline score, group, sex, age, and psychiatric medication) and combined with Rubin's rules and Barnard–Rubin degrees of freedom (Barnard & Rubin, 1999; Rubin, 1987); (d) baseline observation carried forward for non-completers; (e) adjustment for sex, age, psychiatric medication, and co-occurring mental health conditions; and (f) for

lnRMSSD, restriction to recordings rated as good quality at both time points. For the PHQ-9, lnRMSSD, and PSQI, the interaction was also estimated in a Bayesian framework by combining a normal approximation to the likelihood with a skeptical normal prior centered on zero, with a standard deviation equal to half the baseline standard deviation (which places 95% of the prior mass within ± 1 baseline standard deviation). Bayes factors were approximated from the Bayesian information criterion (Wagenmakers, 2007). PHQ-9 response was defined as a reduction of 3 points or more.

Because autistic participants started with lower HRV, an exploratory model tested whether baseline lnRMSSD moderated the VR effect (group × centered baseline score) and whether adding this term changed the group × population estimate. The same check was run for the PHQ-9.

Retention was compared with Fisher's exact test and reported with Wilson 95% confidence intervals. Baseline differences between the samples were tested with Welch *t* tests, Mann–Whitney tests, and Fisher's exact tests, and continuous differences were also adjusted for sex, age, and psychiatric medication. Analyses were run in Python 3.13 with statsmodels 0.15 and SciPy 1.18.

Results

Baseline Characteristics

Table 1 shows baseline characteristics by trial and arm. The two samples were similar in age (non-autistic $M = 21.2$ years, $SD = 1.5$; autistic $M = 21.3$, $SD = 2.7$) and in the proportion of men (40% and 50%). Autistic participants more often reported a co-occurring mental health condition (60% vs. 27%, $p = .038$) and psychiatric medication use (75% vs. 10%, $p < .001$). They had lower resting lnRMSSD (3.47 vs. 4.00; $g = -0.82$, $p = .011$), a difference that remained after adjustment for sex, age, and medication ($p = .048$) and after removal of one implausibly low recording ($g = -0.74$, $p = .016$). They also reported about 110 more minutes of sitting per day ($p = .034$) and had more visceral fat (10.1 vs. 6.2, $p = .049$); visceral fat, BMI, and body fat percentage were all higher after adjustment (adjusted $p = .030$, .039, and .016). Depressive symptoms were somewhat higher in the autism trial (PHQ-9 8.2 vs. 6.2; Mann–Whitney $p = .043$), but the difference disappeared after adjustment for medication ($p = .719$). Five of 20 autistic participants and 4 of 30 non-autistic participants scored 10 or above on the PHQ-9 at baseline. Within the autism trial, participants randomized to VR boxing had higher BMI, body fat, and visceral fat than controls.

Table 1

Baseline Characteristics by Trial and Arm

Characteristic	Non-autistic VR (n = 16)	Non-autistic control (n = 14)	Autism VR (n = 11)	Autism control (n = 9)	p[a]	Adjusted p[b]
Age, years	21.4 (1.5)	21.0 (1.6)	22.2 (3.1)	20.2 (1.9)	.921	—
Men, n (%)	5 (31%)	7 (50%)	6 (55%)	4 (44%)	.567	—
White, n (%)	10 (63%)	7 (50%)	11 (100%)	6 (67%)	.062	—
Co-occurring mental health condition, n (%)	5 (31%)	3 (21%)	6 (55%)	6 (67%)	.038	—
Psychiatric medication, n (%)	1 (6%)	2 (14%)	9 (82%)	6 (67%)	< .001	—
Moderate or extensive VR experience, n (%)	4 (25%)	2 (14%)	2 (18%)	2 (22%)	1.00	—
PHQ-9	6.6 (6.0)	5.6 (4.3)	8.8 (4.2)	7.3 (4.0)	.140	.719
PSQI	6.2 (3.8)	6.7 (3.2)	7.6 (4.3)	6.6 (2.4)	.482	.501
lnRMSSD (ln ms)	4.07 (0.53)	3.92 (0.54)	3.39 (1.00)	3.58 (0.27)	.011	.048
Resting heart rate (bpm)	78.3 (13.7)	80.8 (11.0)	82.9 (14.5)	89.7 (13.1)	.100	.368
BMI (kg/m²)	23.6 (3.2)	22.0 (3.2)	30.0 (6.3)	21.1 (2.3)	.087	.039
Body fat (%)	24.9 (8.4)	21.1 (7.8)	37.6 (12.6)	20.8 (9.1)	.074	.016
Skeletal muscle index (kg/m²)	7.2 (0.9)	7.1 (1.1)	7.5 (0.8)	6.8 (0.9)	.953	.704
Visceral fat level	7.1 (4.1)	5.1 (3.1)	15.1 (7.4)	4.6 (2.5)	.049	.030
Physical activity (MET-min/week)	5909 (5658)	5231 (3552)	3423 (7370)	2895 (2631)	.123	.230
Sitting time (min/day)	291 (124)	399 (157)	433 (174)	472 (208)	.034	.107

Note. Values are *M* (*SD*) unless noted. VR = virtual reality; PHQ-9 = Patient Health Questionnaire-9; PSQI = Pittsburgh Sleep Quality Index; lnRMSSD = natural log of the root mean square of successive differences. lnRMSSD was missing for one non-autistic control, and body composition for one autistic VR participant.

[a] Autistic vs. non-autistic participants: Welch *t* test for continuous and Fisher's exact test for categorical variables. Mann–Whitney $p = .043$ for the PHQ-9 and .001 for physical activity.

[b] Coefficient for autism trial adjusted for sex, age, and psychiatric medication.

Retention, Adherence, and Exercise Dose

All 30 non-autistic participants completed post-testing. In the autism trial, 6 of 11 participants assigned to VR boxing (55%, 95% CI [28, 79]) and 8 of 9 controls (89%) did so. Retention in the VR arm differed between trials (100% vs. 55%, Fisher $p = .006$), whereas retention in the control arms did not (100% vs. 89%, $p = .39$). Four of the five autistic non-completers had symptom data from the first VR session and no data from any later session, so dropout clustered immediately after the first exposure. Non-completers had higher baseline PHQ-9 scores than completers ($M = 10.2$ vs. 7.7) and lower baseline lnRMSSD (3.10 vs. 3.63; 3.55 without the implausible recording).

Among completers, adherence was high in both trials: non-autistic participants attended 5.4 of 6 sessions (91%) and autistic participants 7.5 of 8 (94%). Total VR exposure was similar (209 vs. 254 min; about 38 and 34 min per session), although the schedules differed (weekly vs. twice weekly). In the non-autistic trial, mean session heart rate was 72% of the band-estimated maximum. In the autism trial, 24 of 45 completed sessions were played at the lowest difficulty setting, 14 at the middle setting, and 7 at the highest. Three of the six autistic completers exercised at home.

Depressive Symptoms

Table 2 and Fig. 2 present the trial-specific effects and their difference. The baseline-adjusted VR − control difference in PHQ-9 scores was − 1.99 points (95% CI [− 4.36, 0.39], $p = .097$) in the non-autistic trial, which reproduces the primary report, and − 2.19 points (95% CI [− 6.57, 2.19], $p = .294$) in the autism trial. The difference between trials was − 0.21 points (95% CI [− 4.68, 4.26], $p = .926$). Across sensitivity analyses, the interaction ranged from − 0.21 to + 0.11 (Table 3). Under intention-to-treat imputation, the effect within the autism trial was − 1.88 points (95% CI [− 5.61, 1.86]). The Bayesian posterior for the interaction was centered at − 0.11 (95% credible interval [− 3.32, 3.09]), with a posterior probability of .53 that the effect was larger in the autism trial and a Bayes factor (BF_{01}) of 6.6 in favor of no difference. The intervals were wide, however, and the minimum detectable interaction was about 6.2 points. A reduction of 3 points or more occurred in 7 of 16 VR participants and 2 of 14 controls in the non-autistic trial, and in 4 of 6 VR participants and 2 of 8 controls in the autism trial (Fig. 3a).

Table 2

Baseline-Adjusted Effects of VR Boxing Within Each Trial and the Difference Between Trials

Outcome	Non-autistic trial: VR − control [95% CI]	Autism trial: VR − control [95% CI]	Difference (autism − non-autistic) [95% CI]	p	p_{Holm}
Primary outcome					
PHQ-9	−1.99 [− 4.36, 0.39]	−2.19 [− 6.57, 2.19]	−0.21 [− 4.68, 4.26]	.926	1.00
Secondary outcomes					
lnRMSSD	−0.24 [− 0.73, 0.26]	+ 0.29 [− 0.27, 0.86]	+ 0.53 [− 0.25, 1.32]	.176	1.00
PSQI	−0.23 [− 2.64, 2.17]	−1.23 [− 2.71, 0.26]	−0.99 [− 4.68, 2.69]	.588	1.00
Resting heart rate (bpm)	+ 3.2 [− 6.8, 13.3]	−1.0 [− 15.6, 13.5]	−4.3 [− 22.4, 13.8]	.634	1.00
Exploratory outcomes					
Body fat (%)	+ 1.53 [0.20, 2.85]	+ 1.30 [− 0.84, 3.44]	−0.23 [− 2.77, 2.32]	.857	1.00
Skeletal muscle index (kg/m²)	−0.01 [− 0.15, 0.14]	+ 0.03 [− 0.15, 0.21]	+ 0.04 [− 0.22, 0.29]	.774	1.00
Visceral fat level	+ 0.68 [− 0.03, 1.38]	+ 0.43 [− 0.62, 1.48]	−0.25 [− 1.64, 1.14]	.718	1.00
BMI (kg/m²)	+ 0.42 [− 0.05, 0.89]	+ 0.34 [− 0.29, 0.96]	−0.08 [− 0.99, 0.83]	.854	1.00
Physical activity (MET-min/week)	+ 1209 [− 715, 3132]	+ 2383 [− 318, 5083]	+ 1174 [− 2109, 4457]	.474	1.00
Sitting time (min/day)	+ 86 [− 17, 190]	−201 [− 378, − 23]	−287 [− 470, − 104]	.003	.030

Note. Complete-case analysis of covariance with the post-test score as outcome and the baseline score as covariate. The difference is the group × population coefficient from the pooled model with trial-specific intercepts and baseline slopes. Holm correction applied across the ten interaction tests. *n* (VR/control) = 16/14 in the non-autistic trial (16/12 for lnRMSSD and resting heart rate) and 6/8 in the autism trial (5/8 for body composition). Minimum detectable interaction at 80% power: PHQ-9, 6.2 points; lnRMSSD, 1.08; PSQI, 5.1 points.

Autonomic and Sleep Outcomes

In the non-autistic trial, lnRMSSD was slightly lower after VR boxing than after control (− 0.24, 95% CI [− 0.73, 0.26]); in the autism trial it was higher (+ 0.29, 95% CI [− 0.27, 0.86]). The interaction (+ 0.53, 95% CI [− 0.25, 1.32], p = .176) was not statistically significant, but it kept the same sign in every sensitivity analysis (+ 0.42 to + 0.69; Table 3), and the posterior probability of a positive interaction was .82. Part of

this difference came from the control arms, where lnRMSSD rose slightly in the non-autistic trial (+ 0.10) and fell in the autism trial (− 0.21; Fig. 3b). Sleep quality improved more in the autism trial (− 1.23 vs. −0.23 PSQI points), but the interaction was imprecise (− 0.99, 95% CI [− 4.68, 2.69]). Resting heart rate did not differ between trials (Table 2).

Table 3

Sensitivity of the Group × Population Interaction for Depressive Symptoms and Resting HRV

Analysis	**PHQ-9: estimate [95% CI]**	**p**	**lnRMSSD: estimate [95% CI]**	**p**
Complete case (primary)	−0.21 [− 4.68, 4.26]	.926	+ 0.53 [− 0.25, 1.32]	.176
HC3 robust standard errors	−0.21 [− 5.44, 5.03]	.938	+ 0.53 [− 0.21, 1.27]	.158
Two-stage (trial-specific estimates differenced)	−0.21 [− 5.03, 4.61]	.929	+ 0.53 [− 0.19, 1.25]	.140
Multiple imputation, intention to treat	+ 0.11 [− 4.04, 4.26]	.958	+ 0.46 [− 0.23, 1.15]	.186
Baseline observation carried forward	+ 0.03 [− 4.03, 4.09]	.987	+ 0.42 [− 0.25, 1.10]	.213
Adjusted for sex, age, medication, mental health condition	−0.14 [− 5.65, 5.36]	.958	+ 0.69 [− 0.27, 1.65]	.151
Good-quality HRV recordings only (n = 35)	—	—	+ 0.44 [− 0.36, 1.24]	.272
Bayesian, skeptical prior: posterior mean [95% CrI]	−0.11 [− 3.32, 3.09]	BF_{01} = 6.6	+ 0.23 [− 0.27, 0.73]	BF_{01} = 2.2

Note. Estimates are the difference in VR effects (autism trial minus non-autistic trial). Multiple imputation used chained equations with predictive mean matching (200 imputations) within each trial. The Bayesian analysis used a normal prior centered on zero with a standard deviation of half the baseline SD; BF_{01} > 3 indicates moderate evidence for no difference. CrI = credible interval.

Exploratory Outcomes

Body composition and physical activity did not differ between trials (Table 2). Sitting time was the exception. Autistic participants who did VR boxing reported less daily sitting than controls at post-test (− 201 min/day, 95% CI [− 378, − 23]), whereas non-autistic participants reported more (+ 86 min/day, 95% CI [− 17, 190]). The interaction (− 287 min/day, 95% CI [− 470, − 104]) remained significant after Holm correction (p = .030) and held with robust standard errors (p = .010), with baseline carried forward (− 244 min/day, p = .002), and after exclusion of one implausible value in the autism control arm (− 316 min/day, p < .001; Online Resource 1). Because sitting time was self-reported and the VR arm of the autism trial contributed six participants, this result is best treated as a hypothesis.

Baseline Level or Population?

Baseline lnRMSSD did not moderate the effect of VR boxing on lnRMSSD (group × baseline = − 0.19, 95% CI [− 0.93, 0.56], p = .614), and adding this term barely changed the group × population estimate (+ 0.51, 95% CI [− 0.32, 1.35]). The same held for the PHQ-9 (group × baseline = − 0.06, p = .801). The differences between trials were therefore not explained by where participants started (Online Resource 1).

Discussion

This pooled analysis of two pilot trials asked whether autistic young adults respond to VR boxing the way non-autistic peers do. For depressive symptoms, these data suggest that the effect among people who completed the program was about the same size in both groups, roughly 2 PHQ-9 points relative to no exercise. The more consequential difference was retention. Every non-autistic participant assigned to VR boxing finished the study, while close to half of the autistic participants left, most of them right after the first session.

A 2-point difference on the PHQ-9 is modest, and the confidence intervals in both trials include zero. The estimate is nonetheless in the range reported for exercise in adults with mild depressive symptoms (Noetel et al., 2024; Schuch et al., 2016), and the agreement between trials was close: the point estimates differed by a fifth of a point, and every sensitivity analysis kept the difference near zero. The data do not establish equivalence, since an interaction of 4 points in either direction is still compatible with the results. What they do argue against is the assumption that autistic adults gain less from structured exercise. For autistic participants who continued past the first session, the program worked at least as well as it did for their non-autistic peers.

The retention gap is the more robust finding. It was large (45 percentage points), it was confined to the VR arm, and it was concentrated at the start. Interest does not explain it. In the screening survey that preceded the autism trial, autistic and non-autistic students expressed similar interest in VR boxing (Lee, Vazou, Kim, Ratan, et al., 2026). Something about the first session appears to have been harder for autistic participants, and several features of that session are plausible candidates. A first session combines an unfamiliar headset, a new physical task, fast-moving visual targets, music, and an unfamiliar setting, which together make a dense sensory and social load. Simulator sickness is another candidate. Among completers of the autism trial, lower resting HRV predicted symptoms that persisted across sessions (Lee, in press), and in the present data non-completers also had lower baseline HRV than completers. Non-completers also started with more depressive symptoms, which may have made a demanding first session harder to repeat. These explanations are not mutually exclusive, and the present data cannot separate them.

The practical implication is that the first session deserves as much design attention as the exercise dose. Approaches worth testing include a separate familiarization visit without exercise demands, a short first session at the lowest intensity, a video preview of the session, a choice of exercise location

from the outset, and resting HRV screening to identify people who may need a slower introduction rather than exclusion (Lee, in press). In the next trial, retention should be a primary outcome rather than a byproduct.

The HRV results extend the affective–autonomic dissociation reported in the non-autistic trial (Lee, Vazou, Kim, & Hauck, 2026). Non-autistic participants improved in mood without a corresponding HRV change, whereas autistic participants showed a small HRV increase alongside the mood change. Baseline HRV did not account for the difference, which argues against a simple ceiling or regression-to-the-mean explanation. Two other explanations seem more plausible. The autism trial delivered sessions twice a week, and the non-autistic trial report noted that one session per week is probably too infrequent to change resting HRV. In addition, about half of the gap came from the control arms moving in opposite directions, which may reflect time-varying influences such as the academic calendar acting on two small samples. A trial that assigns both populations to the same schedule is needed before this can be read as a population difference.

The sitting-time result is harder to interpret. Autistic participants who did VR boxing reported sitting more than 3 hours less per day than autistic controls at post-test after baseline adjustment, whereas non-autistic participants reported sitting more. One reading is that a structured, twice-weekly routine displaced sedentary time for autistic participants, who started with more of it. Another is measurement error in a self-reported item with large variance. Device-based measurement would resolve this.

Several limitations apply. First, population and trial were the same variable. Each trial's effect is protected by randomization, but the comparison between them is observational: the trials differed in session frequency and duration, the option to exercise at home, and the medication profile of their participants (10% vs. 75% on psychiatric medication), and any of these could produce or hide a population difference. Second, the samples were small. Only six autistic participants completed VR boxing, so every estimate involving that group is imprecise, and interactions require far larger samples than main effects (Brookes et al., 2004). Small samples also allowed chance imbalance within the autism trial, whose VR arm had higher BMI and visceral fat than its control arm. Third, neither trial was prospectively registered, and the pooled comparison was planned after data collection. The analysis plan was kept simple and all outcomes are reported to limit selective reporting, but the results should be treated as exploratory. Fourth, the complete-case analysis assumes that autistic completers are representative of everyone randomized. The imputation and baseline-carried-forward analyses gave similar results, but both rest on assumptions about why participants left. Fifth, autism diagnosis was reported by participants or caregivers rather than confirmed, and both samples were university students, which limits generalizability to autistic adults with higher support needs. Finally, physical activity and sitting time were self-reported.

The next step is a single trial that randomizes autistic and non-autistic adults to the same VR boxing protocol, registers its analysis plan in advance, and treats retention as a co-primary outcome. Building an

onboarding component into the intervention and measuring sedentary time with devices would address the two open questions raised here.

Conclusion

Autistic young adults who completed VR boxing showed a reduction in depressive symptoms similar to that of non-autistic peers. The difference between the groups lay in who stayed: nearly half of the autistic participants left after the first session. For exercise programs aimed at autistic adults, getting people through the first session may matter as much as what happens after it.

Declarations

Author Contribution

Y.L. conceived the pooled analysis, contributed to the design and conduct of both trials and to data collection, performed all analyses, and wrote and revised the manuscript.

Acknowledgement

The author thanks the participants and their families for their time and contribution to this research.

Data Availability

The de-identified participant-level data analyzed in this study are not publicly available because participants did not consent to public data sharing and the small sample of autistic participants increases the risk of re-identification. The data and analysis code are available from the corresponding author on reasonable request.

References

1. Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955. https://doi.org/10.1093/biomet/86.4.948
2. Bradley, R., & Newbutt, N. (2018). Autism and virtual reality head-mounted displays: A state of the art systematic review. *Journal of Enabling Technologies*, *12*(3), 101–113. https://doi.org/10.1108/JET-01-2018-0004
3. Brookes, S. T., Whitely, E., Egger, M., Davey Smith, G., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology*, *57*(3), 229–236. https://doi.org/10.1016/j.jclinepi.2003.08.009

4. Bull, F. C., Maslin, T. S., & Armstrong, T. (2009). Global Physical Activity Questionnaire (GPAQ): Nine country reliability and validity study. *Journal of Physical Activity and Health*, *6*(6), 790–804. https://doi.org/10.1123/jpah.6.6.790
5. Buysse, D. J., Reynolds, C. F., Monk, I. I. I., Berman, T. H., S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4
6. Cheng, Y. C., Huang, Y. C., & Huang, W. L. (2020). Heart rate variability in individuals with autism spectrum disorders: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *118*, 463–471. https://doi.org/10.1016/j.neubiorev.2020.08.007
7. Croen, L. A., Zerbo, O., Qian, Y., Massolo, M. L., Rich, S., Sidney, S., & Kripke, C. (2015). The health status of adults on the autism spectrum. *Autism*, *19*(7), 814–823. https://doi.org/10.1177/1362361315577517
8. Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., Lancaster, G. A., & PAFS consensus group. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Bmj*, *355*. https://doi.org/10.1136/bmj.i5239. Article i5239.
9. Hillier, A., Buckingham, A., & Schena, D., II (2020). Physical activity among adults with autism: Participation, attitudes, and barriers. *Perceptual and Motor Skills*, *127*(5), 874–890. https://doi.org/10.1177/0031512520927560
10. Hollocks, M. J., Lerh, J. W., Magiati, I., Meiser-Stedman, R., & Brugha, T. S. (2019). Anxiety and depression in adults with autism spectrum disorder: A systematic review and meta-analysis. *Psychological Medicine*, *49*(4), 559–572. https://doi.org/10.1017/S0033291718002283
11. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.
12. Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
13. Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
14. Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, *8*, 213. https://doi.org/10.3389/fpsyg.2017.00213
15. Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819–829. https://doi.org/10.1016/S2215-0366(19)30289-5
16. Lee, Y. (in press). *Resting heart rate variability and the persistence of simulator sickness during virtual reality exercise in young adults with autism spectrum disorder: A proof-of-principle study*. Research in

Autism.

17. Lee, Y., Vazou, S., Kim, S., & Hauck, J. (2026). *Effects of a 6-week virtual boxing on depressive symptoms and heart rate variability in adults: A pilot randomized controlled trial* [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-9707388/v1
18. Lee, Y., Vazou, S., Kim, S., Ratan, R., & Hauck, J. (2026). *Interest in immersive virtual reality boxing exercise among college students with and without autism spectrum disorder* [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-9246183/v1
19. Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008
20. Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549
21. Must, A., Phillips, S., Curtin, C., & Bandini, L. G. (2015). Barriers to physical activity in children with autism spectrum disorders: Relationship to physical activity and screen time. *Journal of Physical Activity and Health*, *12*(4), 529–534. https://doi.org/10.1123/jpah.2013-0271
22. Noetel, M., Sanders, T., Gallardo-Gómez, D., Taylor, P., del Pozo Cruz, B., van den Hoek, D., Smith, J. J., Mahoney, J., Spathis, J., Moresi, M., Pagano, R., Pagano, L., Vasconcellos, R., Arnott, H., Varley, B., Parker, P., Biddle, S., & Lonsdale, C. (2024). Effect of exercise for depression: Systematic review and network meta-analysis of randomised controlled trials. *Bmj*, *384*. https://doi.org/10.1136/bmj-2023-075847. Article e075847.
23. Obrusnikova, I., & Cavalier, A. R. (2011). Perceived barriers and facilitators of participation in after-school physical activity by children with autism spectrum disorders. *Journal of Developmental and Physical Disabilities*, *23*(3), 195–211. https://doi.org/10.1007/s10882-010-9215-z
24. Riley, R. D., Debray, T. P. A., Fisher, D., Hattle, M., Marlin, N., Hoogland, J., Gueyffier, F., Staessen, J. A., Wang, J., Moons, K. G. M., Reitsma, J. B., & Ensor, J. (2020). Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Statistics in Medicine*, *39*(15), 2115–2137. https://doi.org/10.1002/sim.8516
25. Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696
26. Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric Research*, *77*, 42–51. https://doi.org/10.1016/j.jpsychires.2016.02.023
27. Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*(3), 201–216. https://doi.org/10.1016/S0165-0327(00)00338-4

28. Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105
29. Wang, Z., Song, B., Liu, C., Ma, H., Bai, Z., Carneiro, M. A. S., Youssef, L., Chen, C., Zhang, L., Wang, D., & Wang, D. (2025). Effects of boxing exercise in people with Parkinson's disease: A systematic review. *Frontiers in Aging Neuroscience*, *17*, 1505326. https://doi.org/10.3389/fnagi.2025.1505326

Figures

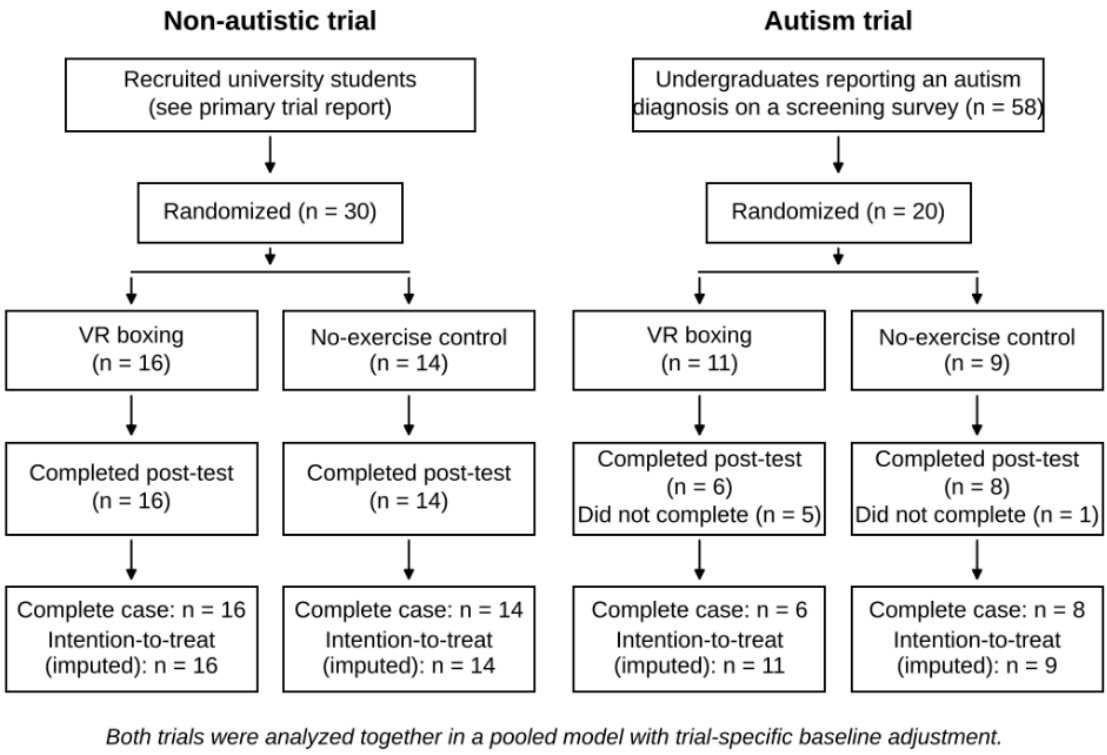


Figure 1

Participant flow in the two pilot trials

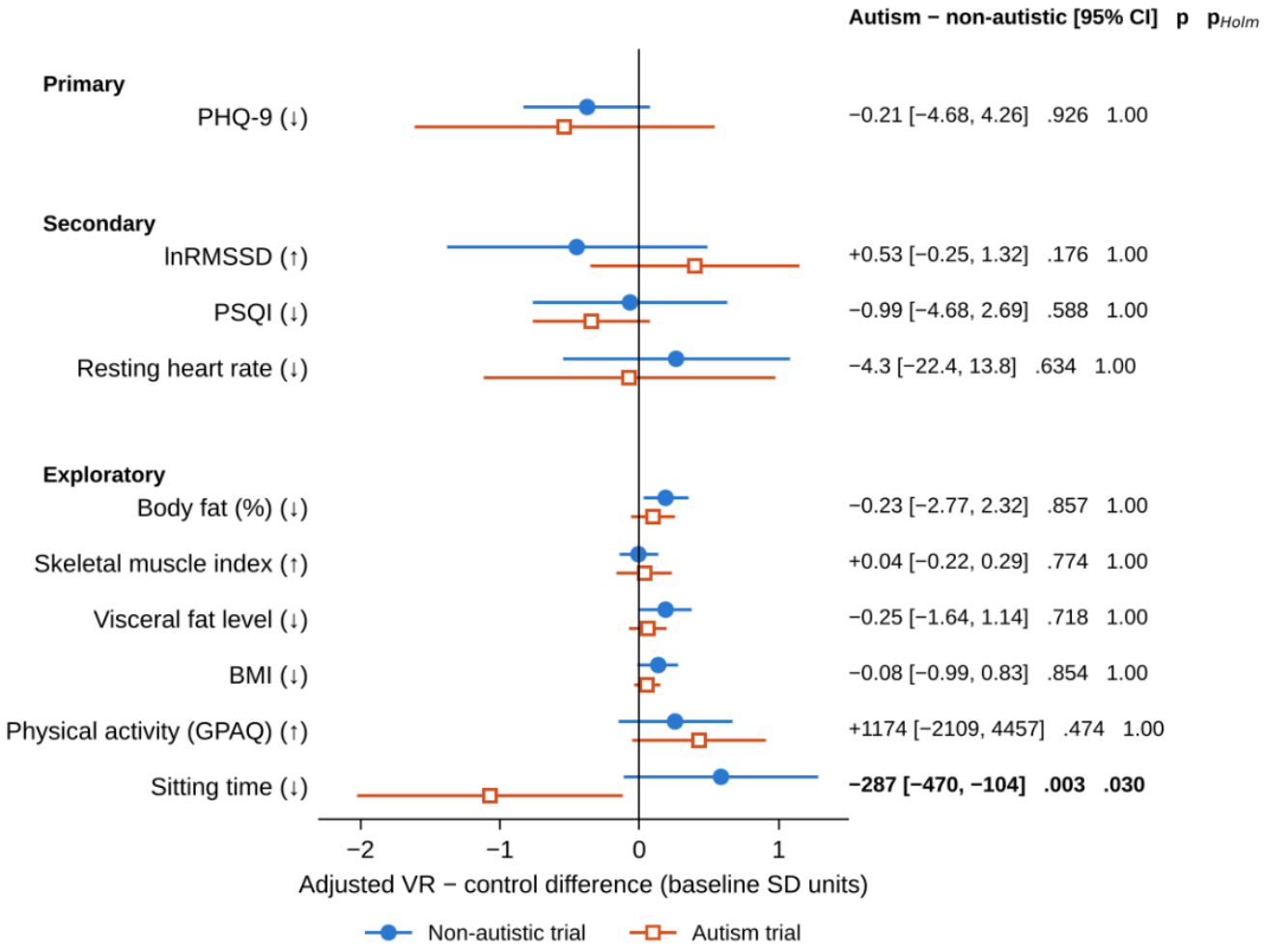


Figure 2

Baseline-adjusted effects of VR boxing relative to control in each trial. Points show the adjusted difference divided by the trial's baseline standard deviation, and horizontal lines show 95% confidence intervals (circles, non-autistic trial; squares, autism trial). Arrows after outcome names give the direction of improvement. The right-hand column gives the difference between trials in original units with unadjusted and Holm-adjusted *p* values

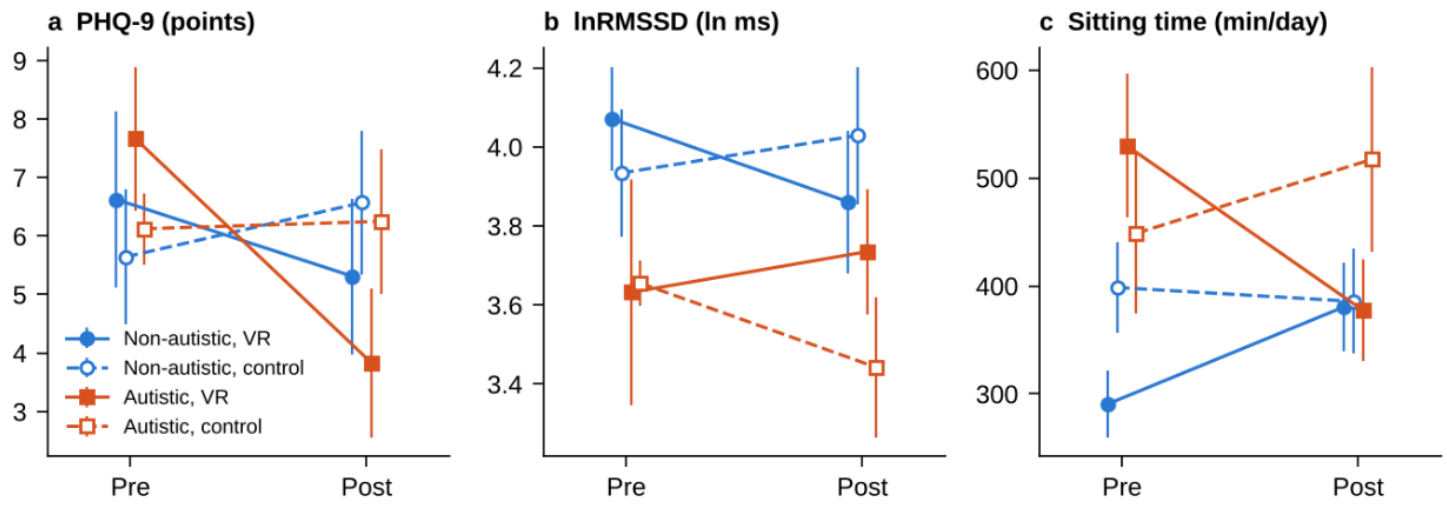


Figure 3

Pre- and post-test means (±SE) for completers by trial and arm: **a** PHQ-9, **b** lnRMSSD, **c**sitting time. Solid lines with filled markers show VR boxing and dashed lines with open markers show control; circles denote the non-autistic trial and squares the autism trial

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf