Supplementary Online Content

Boyne P, Billinger SA, Reisman DS, et al. Optimal intensity and duration of walking rehabilitation in patients with chronic stroke: a randomized clinical trial. *JAMA Neurol*. Published online February 23, 2023. doi:10.1001/jamaneurol.2023.0033

eAppendix
eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.

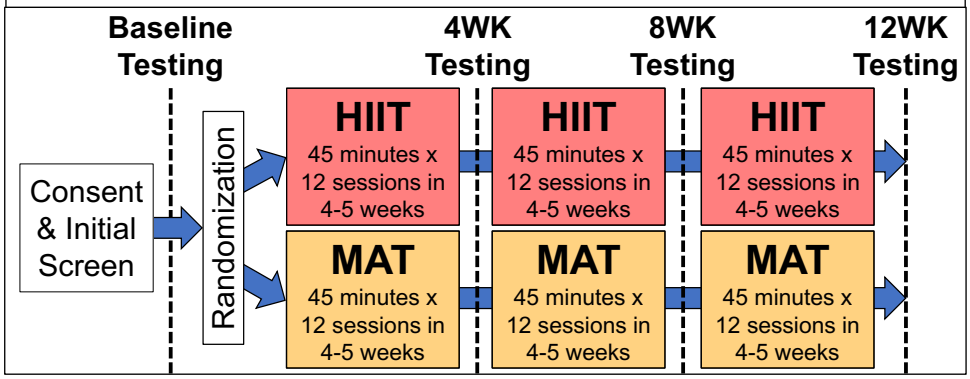
eAppendix

Supplementary Methods

Study design

Eligible participants were randomized to either HIIT or MAT, with a target training volume of 45 minutes, 3 times per week for 12 weeks. This was divided into three 4-week (12-session) blocks separated by outcome testing (Figure S1). In each training block, up to one additional week was allowed to make up for missed sessions if needed. Outcomes were assessed by blinded raters before randomization (Baseline) and after 4, 8 and 12 weeks of training (4WK, 8WK, 12WK).

Figure S1. Study design
HIIT, high-intensity interval training; MAT, moderate-intensity aerobic training; WK, week



Multisite standardization

Procedures were standardized across sites using a detailed manual of procedures, video-based online personnel training, 2-day in-person training at each site and monthly web-meetings. This study also used direct-electronic data entry into REDCap forms that were programmed to provide real-time protocol reminders, automated calculations, prompts and feedback during each study visit. In addition, a software application used during each treatment visit was programmed to prompt training start/stop times and provide real-time intensity feedback.

Interventions

Training sessions were directed by a physical therapist with support from a research assistant as needed for data collection. Habitual orthotic devices were used and occasionally supplemented by additional orthotics to protect against ankle inversion sprain or severe knee hyperextension during training, based on the judgment of the treating therapist. Participants were not physically assisted with stepping but were guarded as needed and assisted for injury prevention during instances of severe gait instability or loss of balance. Participants who enrolled after the beginning of the COVID-19 pandemic wore personal protective equipment during testing and treatment sessions according to the infection control protocols at each site.

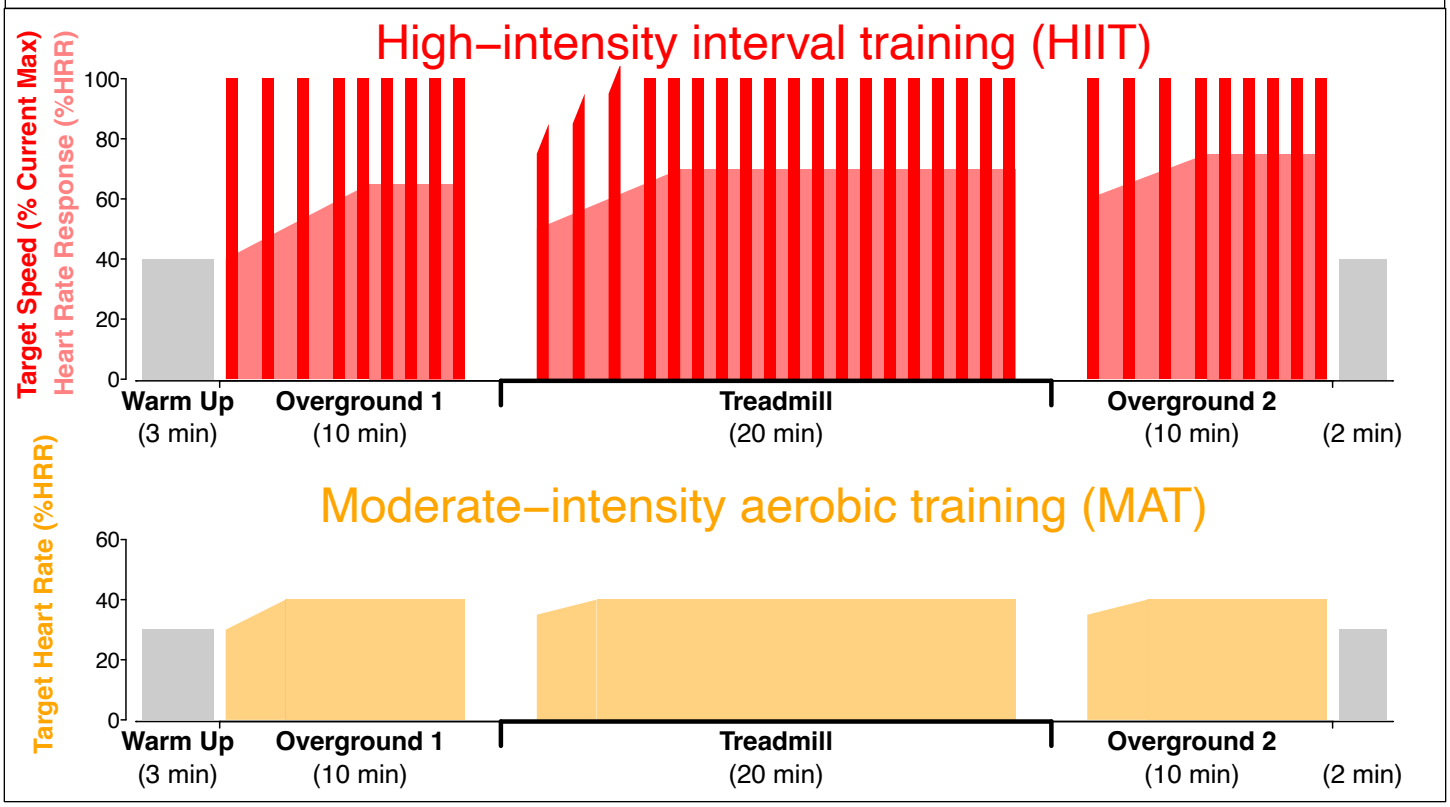
During overground training bouts, participants walked back and forth in a corridor. Habitual assistive devices were used during initial bouts, then therapists attempted to progress the participant to less restrictive device(s) if it better enabled achievement of the target intensity. When training on a treadmill (LiteGait Gaitkeeper 2200T, Mobility Research, Tempe, AZ, USA; or L8 Rehab Treadmill, Landice, Randolph, NJ, USA), participants wore a fall protection harness without any weight support (Balance Harness, MASS Rehab, Dayton, OH) connected to an overhead system (RehabMODULE, MASS Rehab, Dayton, OH, USA) and used a handrail connected to the fall protection system for balance support.[1] This handrail was individually height-adjusted during treadmill training to allow upright walking posture.

Training intensity was monitored and recorded using Bluetooth Smart heart rate (HR) transmitters (H7 and H10, Polar Electro Oy, Kempele, Finland; or Rhythm24, Scosche Industries, Inc., Oxnard, CA, USA) wirelessly connected to an iPod touch (Apple, Cupertino, CA, USA) using the iCardio application (FitDigits, Inc., Ventura, CA, USA).[2, 3] The training protocols and individual prescribed target HR zones were programmed into the application to time each training bout, signal burst/recovery transitions (for HIIT) and provide real-time intensity guidance. A forearm-worn iPod mount enabled hands-free monitoring for the treating therapist while guarding the participant.

Training HR zones were calculated using the HR reserve (HRR) method: (HR_{peak} – HR_{rest}) x % HRR target + HR_{rest}.[4] During training, HR_{peak} was taken as the highest instantaneous HR obtained in any prior exercise testing session for that participant. HR_{rest} was measured at the beginning of each session in both sitting and standing. For this study, the HR_{rest} value in standing was used for target HR calculations. This was based on our pilot testing in which some participants with more severe deconditioning were found to have large

HR increases just from standing that would have otherwise exceeded some of the target HR ranges and precluded walking practice if we had used the true HR_{rest} value for target HR calculation.
Each session, the training protocol for both groups included: 1) a 3-minute warm-up of overground walking at a speed that increased HR to 30-40% HR reserve (HRR); 2) a 10-minute bout of overground HIIT or MAT; 3) a 20-minute bout of treadmill HIIT or MAT; 4) another 10-minute bout of overground HIIT or MAT; and 5) a 2-minute cool down at 30-40% HRR (Figure S2).

Figure S2. Training Protocol Schematics. HIIT (upper panel) uses 30 second bursts at 100% maximum safe speed (dark red bars) alternated with 30-60 second passive recovery periods and within-session speed progression as able (not shown). This typically elicits a mean aerobic intensity >60% heart rate reserve (HRR; light red shading). The first three treadmill bursts depict example within-burst speed increases to find the current maximum safe treadmill speed. MAT (lower panel) uses continuous walking with speed adjusted as needed to maintain the moderate aerobic intensity target, which starts at 40% HRR (shown) and progresses by 5% HRR every 2 weeks, up to 60% HRR (not shown).



High-intensity interval training (HIIT) protocol

The HIIT group used a 'short-interval' HIIT protocol that was specifically developed for locomotor exercise post-stroke.[2, 3, 5, 6] It involved repeated 30 second bursts of walking at maximum safe speed, alternated with 30-60 second passive recovery periods, targeting an average aerobic intensity above 60% HRR. During overground HIIT, burst speed was maximized by placing cones or beanbags on the floor to mark the distance covered during each burst and encouraging participants to continually work towards making it past the marker before time ran out. During treadmill HIIT, burst speed was selected to provide the maximum safe challenge for that participant and was progressed as able (or regressed as needed) throughout each session based on performance criteria to maintain constant challenge.[3]

Moderate-intensity aerobic training (MAT) protocol

Following current best-practice guidelines for stroke rehabilitation,[7, 8] the MAT group performed continuous walking practice with speed adjusted to maintain an initial target HR of 40 ± 5% HRR, progressing by 5% HRR every 2 weeks, up to 60% HRR as tolerated.[1]

Treatment fidelity data collection

- *Training volume* was measured by the number of sessions attended, the number of treadmill and overground training bouts started and the total number of training minutes performed.
- *Neuromotor intensity* was measured by peak training speed in each bout. Overground speeds were captured using a stopwatch at the beginning and end of each overground training bout as the participants

crossed over markings for the 10-meter walk test in the overground training corridor. Treadmill speeds were recorded from the treadmill display.

- *Aerobic intensity* was measured by HR data collected continuously by the Bluetooth HR monitors and iPod touch iCardio application during each training bout. These data were exported at 1/3 Hz and processed to calculate mean steady-state HR (excluding the first 3 minutes) and max HR for each bout as %HRR.
  - For real-time treatment monitoring, %HRR values were calculated using standing HR_{rest} and the most current HR_{peak}, as described above. However, the use of standing HR_{rest} makes such %HRR values incomparable to previous studies and allowing HR_{peak} updates at 4WK or 8WK could distort %HRR comparisons across sessions and/or treatment groups. Therefore, to facilitate these comparisons, %HRR values were also calculated using true HR_{rest} (the smaller of the two HR_{rest} values obtained from each session, which was typically the sitting HR_{rest}) and HR_{peak} from baseline testing only.
  - HR data preprocessing included spike-filtering and plotting to visually check for invalid data (signal dropout, residual non-physiologic spikes, episodes of supraventricular tachycardia and large elevations clearly attributed to coughing or laughing), while referencing notes and other data from each bout. Where possible, invalid data segments were filled using linear interpolation. When a bout began with invalid data, the first value at the beginning of the adjacent bout(s) in the same session was used to initialize the interpolation, to approximate the typical HR increase from a low starting value at the beginning of the bout. When a bout ended with invalid data, the final valid data point was carried forward, to approximate the typical steady mean trajectory at the end of the bout. However, when this did not reasonably approximate the likely mean HR trajectory, the HR data from that bout were discarded as invalid (e.g. when >50% of the HR data from a bout were invalid/missing or the mean trajectory was steeply sloped). If resting HR was missing or invalid for a session (e.g. due to an episode of supraventricular tachycardia that subsided during the session), it was imputed for HRR calculation by taking the mean resting HR across sessions for that participant.
- *Anaerobic intensity* was measured by blood lactate concentration after the treadmill training portion of the middle training session in each training week, using a finger stick and a point-of-care blood lactate analyzer (Lactate Plus, Nova Biomedical, Waltham, MA, USA).
- *Practice repetition* was measured by step count, recorded each session using a Stepwatch Activity Monitor (Modus Health, LLC, Edmonds, WA, USA) on the non-paretic ankle.[9]
- *Rating of perceived exertion* (RPE) for the session was measured by the Borg 6-20 scale[4] at the end of each session.

Outcome assessment

Outcomes were assessed by blinded raters before randomization (Baseline) and after 4, 8 and 12 weeks of training (4WK, 8WK and 12WK). All measures were assessed in a single visit and were arranged in standardized order to mitigate fatigue effects whenever possible, with overground walking tests followed by seated questionnaires then treadmill exercise testing.[10] Testing visits were scheduled 2-7 days after the last treatment visit in the preceding block and subsequent treatment blocks were scheduled to begin within 10 days of the last treatment visit. An individual participant was tested by the same rater at all time points whenever possible. Similar to training, participants used habitual assistive and orthotic devices during testing, wore a fall protection harness during treadmill walking and were not physically assisted but were guarded by a physical therapist and provided injury prevention assistance if needed.

Primary outcome measure

The primary outcome measure was walking capacity, measured by distance walked during the 6-minute walk test (6MWT).[11, 12] The 6MWT was the primary outcome measure because it can be influenced by both gait speed and aerobic fitness,[7, 8] the two primary factors contributing to limited walking capacity post-stroke and the two primary targets of HIIT and MAT.[13] Further, this test explains more variance in home and community ambulation than any other laboratory measure for persons with stroke,[14-16] is a primary characteristic distinguishing between home and community ambulators,[17] and is significantly associated with post-stroke quality of life.[18] The 6MWT has also been accepted by the FDA as a primary outcome for registered trials across various neurologic and cardiopulmonary diagnoses.[19-23] After stroke, it has excellent test-retest reliability (ICC: 0.97-0.99)[24-27] and adequate inter-rater reliability (ICC: 0.78).[28] The minimal clinically important difference is 20 meters and 50 meters is a large improvement.[29] It was standardized according to guidelines,[12] allowing minor common modifications to the testing course at each site.[10]

Secondary outcome measures

- Gait speed at self-selected and fastest speeds, measured by the 10-meter walk test.[1, 30]
- Self-reported fatigue over the past 7 days, measured by the Patient Reported Outcomes Measurement Information System (PROMIS)-Fatigue Scale version 8a.[31, 32]
- Aerobic capacity, measured by oxygen consumption rate (VO_2) at the ventilatory threshold during a treadmill graded exercise test (GXT).[33] The ventilatory threshold is a submaximal transition point that represents the upper intensity limit of prolonged aerobic activity,[34-36] beyond which anaerobic metabolism becomes more dominant, limiting sustainability.[34, 35, 37] Compared with peak VO_2 during an exercise test, which is often confounded by motor impairment after stroke,[33, 34, 37, 38] the ventilatory threshold appears to provide a more specific measure of aerobic capacity in this population.[33]

Treadmill exercise testing

The treadmill GXT was done with ECG monitoring, blood pressure testing and gas exchange analysis, using a metabolic cart (TrueOne 2400, ParvoMedics, Salt Lake City, UT, USA) with a facemask interface. A blinded physical therapist monitored gait safety while additional study team member(s) monitored ECG, blood pressure and metabolic data. The symptom-limited GXT protocol was standardized to be 0.3 mph at 0% grade for the first 3 minutes, then to increase by 0.1 mph every 30 seconds. If a participant reached 3.5 mph, that speed was maintained, and the incline started increasing by 0.5% every 30 seconds. The test continued until the participant requested to stop, drifted backward on the treadmill and was unable to recover, had severe gait instability judged to pose an imminent safety risk, or reached a cardiovascular safety limit.[4] At least 10 minutes after the GXT, participants also attempted a 3-minute verification test without respiratory gas collection to help ensure that the highest possible HR was reached to guide individualized training intensity prescription. This test was done at the peak successful speed and grade from the GXT or at a higher speed/grade if the blinded testing therapist judged that it would be feasible. It was stopped before 3 minutes if one of the GXT stop criteria occurred sooner.

Ventilatory threshold determination

Ventilatory threshold identification from the GXT metabolic data was done using a semi-automated method. First, breath-by-breath data from each test were resampled to 5 second averaging and truncated to the GXT duration, including VO_2, the rate of carbon dioxide production (VCO_2) and the expiratory air flow volume (VE). Excess CO_2 was calculated as $(VCO_2^2 / VO^2) – VCO^2$, with all values in L/min, and the ventilatory equivalents of O_2 and CO_2 were calculated as VE/VO_2 and VE/VCO_2.[39-41] Automated threshold detection was then applied separately to the excess CO_2 and VE/VO_2 time series[39-41] using the R[42] package 'mcp'[43] to identify the posterior distribution likelihood of the ventilatory threshold at each 5 second time point. This was performed separately after smoothing the data using 0, 10 and 20 second rolling averages. The median of the posterior distributions across all methods was taken as an automated point estimate for time to ventilatory threshold.

Two raters then independently verified or adjusted this automated estimate while blinded to group assignment. Raters viewed graphs of VO_2, VCO_2 and VE by time, excess CO_2 by time, VE/VO_2 and VE/VCO_2 by time and VCO_2 by VO2 (V slope), across all testing timepoints for each participant. Automated posterior distribution likelihoods and point estimates were also displayed on the graphs. The ventilatory threshold was identified by an upward break point in the excess CO_2 and VE/VO_2 data, disproportionate to any VE/VCO_2 increase, where the slope of the VCO_2 by VO_2 relationship increased above 1.[39-41] If no threshold occurred during the test, time to threshold was set at the end of the GXT. Inter-rater reliability was excellent ($ICC_{2,1}$: 0.98 [95%CI: 0.97-0.99], mean absolute error: 0.47 $\pm$ 0.87 minutes) despite a small systematic difference in mean time-to-threshold between raters (10.2 $\pm$ 4.9 versus 10.5 $\pm$ 4.7 minutes, $p<0.0001$). Final time-to-threshold values were determined by agreement (10.4 $\pm$ 4.7 minutes) and VO_2 at the ventilatory threshold was calculated by averaging the VO_2 data in a 20 second window around this time point.

Other exercise test measures to facilitate interpretation and exploratory outcomes

- Exercise capacity (not necessarily specific to aerobic fitness) was measured by GXT duration, peak GXT speed, peak GXT VO_2, peak GXT HR and HR_{peak} (highest HR between GXT and verification test). Peak VO2 was calculated from 5 second averaged data after smoothing with a 20 second rolling average. Peak HR was the instantaneous peak from the ECG data.
- To facilitate interpretation of whether maximal *aerobic* capacity was reached during exercise testing:
    - HR_{peak} was also expressed as % age-predicted maximal HR, calculated using 164 – (0.7 x age [years])[44] for participants taking a β-blocker medication and 206.9 – (0.67 x age [years])[45] otherwise. Values are typically $\geq$90% at maximal aerobic capacity.[4]

- Peak respiratory exchange ratio (VCO_2/VO_2) was calculated, using the same preprocessing as peak VO_2. Values are typically at least ≥1.05 at maximal aerobic capacity.[4]
- Rating of perceived exertion (Borg 6-20) was obtained at the end of the GXT. Values are typically ≥17 at maximal aerobic capacity.[4]

- Treadmill speed at the ventilatory threshold was measured to facilitate interpretation of any changes in VO_2 at the ventilatory threshold. This represents the fastest speed that should have prolonged sustainability without accumulating anaerobic metabolites. It was calculated by averaging the protocol speed values in a 20 second window around the ventilatory threshold time.
- Metabolic cost of treadmill gait [mLO_2/kg/m] during the GXT was calculated as VO_2 [mL/kg/min] / speed [m/min], where lower values represent more efficient gait.[46] Values were averaged in the last 3 minutes of each GXT. However, since faster speeds are typically more efficient after stroke,[46] we also averaged the metabolic cost data in the last 3 minutes of the *shortest* test for each participant. This matched the speeds across time points to assess how much any changes in efficiency were related to (or independent from) any changes in speed.[47]
- Gait stability was measured with the functional ambulation category,[48] based on participant performance during the 10-meter and 6-minute walk tests. Since we did not observe participants walking on nonlevel surfaces, the highest category (independent on nonlevel surfaces) was collapsed into the second highest category (independent on level surfaces), yielding a score range from 0-4+.
- Perceived balance confidence was measured with the activities-specific balance confidence (ABC) scale, a questionnaire with 16 items averaged to yield a total confidence score from 0-100%.[49, 50]
- Quality of life related to mobility, self-care, usual activities, pain/discomfort and anxiety/depression were measured with the EuroQOL-5D-5L questionnaire.[51, 52] A total 'misery score' was calculated by summing the 5 item scores.[52] Individual item scores were also analyzed for mobility and usual activities, given that these constructs were targeted by the interventions. Higher values represent worse problems for each of these scores.

Statistical analysis

Analyses followed a prespecified plan,[10] used an *intent-to-treat* approach and were done with SAS®, version 9.4 (SAS Institute, Cary, NC, USA). The primary study statistician (JK) remained blinded to treatment groups until after the primary analysis. To assess randomization performance, baseline participant characteristics were compared between groups with independent t-tests and Fisher exact tests. Baseline data were also expressed as a percentage of normative predicted values for self-selected gait speed,[53] 6MWT distance,[54] VO_2 at the ventilatory threshold[36, 55] and $VO_{2\text{-peak}}$.[55] The HR_{peak} increase attributable to the verification test was quantified by comparing the GXT peak HR to HR_{peak} with a paired t-test.

Treatment fidelity analysis

Training intensity, repetition and perceived exertion data were compared between groups to evaluate treatment fidelity. Each of these dependent variables was tested in a separate statistical model, which included fixed effects for treatment group, session number (modeled as a categorical effect), training bout (for variables collected at overground 1, treadmill and overground 2 each session) and all possible interactions. Repeated sessions within the same participant were modeled with compound symmetry covariance and repeated bouts within the same session were modeled with unconstrained covariance. Variances were not constrained to be the same between sessions or between treatment groups. Missing data were handled with the method of maximum likelihood.

Adverse event (AE) analysis

To assess the relative odds of harms, post-randomization AEs were compared between treatment groups with logistic regression. Separate models were tested for overall AEs and for each AE categorization, using the number of participants with an AE in that category as the dependent variable and fixed effects for [treatment group], [study site] and [baseline walking limitation severity (<0.4, 0.4-1.0 m/s)]. If only one group had AE(s), a continuity correction was added of 0.5 AEs to each group to permit calculation of the odds ratio. In this case, it was not possible to adjust for study site or baseline limitation severity.

Hypothesis testing

Primary hypothesis testing used a linear model with the 6MWT as the dependent variable and fixed effects for treatment group, testing time point (Baseline, 4WK, 8WK, POST), group-by-time interaction, study site, study site-by-time interaction, baseline walking limitation severity and baseline walking limitation severity-by-time interaction, with unconstrained covariance between repeated testing time points within the same

participant and a significance threshold of $p<0.05$. Secondary outcomes were tested using the same model, with false discovery rate (FDR) correction[56] to the significance threshold to control for multiple testing time points (4WKΔ, 8WKΔ, 12WKΔ) and the 5 primary or secondary outcome measures. Other exercise test measures and exploratory outcomes were tested with FDR correction across testing time points and all tested measures. These analyses handled any missing data with the method of maximum likelihood, which assumes that data were missing at random.[57]

Analysis stratified by baseline walking limitation severity

To preliminarily examine whether the primary results differed for participants with severe versus mild/moderate baseline walking limitations, we added a baseline walking limitation severity-by-group-by-time interaction effect to the primary model. In chronic stroke, persons with mild/moderate walking limitations have been found to have better responsiveness to locomotor exercise.[2, 58]

Sensitivity analysis for missing data assumptions

While the 'missing at random' assumption is common in the analysis of clinical trials, it is not testable and could be violated in many typical circumstances, like when outcome data are missing because of AE-related participant withdrawal.[57, 59] To assess how much the results depended on the 'missing at random' assumption, we repeated the primary analysis assuming that participants with missing outcomes due to AE-related withdrawal had poor outcomes, similar to Duncan and colleagues.[60] For any outcome data point missing because of AE-related withdrawal, the true value was assumed to be distributed around either the baseline or the last observation for that participant, whichever was smaller. If the data point missingness was not AE-related, the true value was assumed to be distributed around the last observation for that participant. These distributions for the true values were assumed to be normal with a standard deviation of 15 m, matching the observed standard deviation of 6MWT changes after 4 weeks of no intervention in a similar population.[2] Using the general multiple imputation framework,[57, 59] 50 datasets were generated by random sampling to impute the missing values, each dataset was analyzed, and the model estimates were pooled across datasets.

Sample size

This study was powered to detect a between-group difference of 20 m in 6MWT change[29], using the software 'GLIMMPSE'.[61] We estimated that the MAT group would improve by 15 m every 4 weeks[6] and set the HIIT estimate 20 m larger at each time point. Variance & covariance parameters were estimated by pooling data across our prior 4-week pilot studies and extrapolating the repeated measures correlations involving the later 8WK and 12WK time points using the highest suggested exponential decay rate (0.5).[61] These calculations indicated a total target sample size of 40 for ≥80% power at a two-sided significance threshold of 0.05. To account for up to 20% attrition, we initially planned to enroll and randomize 50 participants. However, after having to withdraw four participants directly due to COVID-19 shutdown, we opted to increase the enrollment target to 55. This decision was made before any analysis of outcome data.

Supplementary Results

Treatment fidelity

In total during the attended treatment sessions, participants initiated 5,006 (~100%) of 5,025 planned overground and treadmill bouts (HIIT, 2,383/2,400 [99%]; MAT, 2,623/2,625 [~100%]), and performed 66,598 (99%) of 67,000 planned training minutes (HIIT, 31,662/32,000 [99%]; MAT, 34,936/35,000 [~100%]), including intermittent rest breaks.

- Peak training speed was recorded during 4,970 (99%) of the 5,006 initiated bouts (HIIT, 2,364/2,383 [99%]; MAT, 2,606/2623 [99%]). Compared with MAT, HIIT involved significantly faster training speed during all bouts and training blocks (Figure 2; Table S1). For HIIT, training speed was significantly faster on the treadmill versus overground. Conversely for MAT, speed was significantly faster overground versus the treadmill. Training speed significantly increased across sessions for both protocols.
- Valid HR recordings were obtained for 65,666 (99%) of the 66,598 performed training minutes (HIIT, 31,211/31,662 [99%]); MAT, 34,456/34,936 [99%]). Compared with MAT, HIIT involved significantly higher mean and max HR during all bouts and training blocks. For both HIIT and MAT, HR was significantly higher on the treadmill vs overground and HR significantly increased across sessions when expressed relative to baseline HR_{peak}.
- Blood lactate was recorded after the treadmill bout in the middle session of 527 (94%) of 563 attended training weeks (HIIT, 246/268 [92%]; MAT, 281/295 [95%]). Compared with MAT, HIIT elicited a significantly higher lactate response overall.
- Step counts were recorded during 1,464 (87%) of the 1,675 attended treatment sessions (HIIT, 712/800 [89%]; MAT, 752/875 [86%]). Compared with MAT, HIIT produced significantly lower step counts during all training blocks. For both HIIT and MAT, step counts significantly progressed across sessions, but significantly less so for HIIT versus MAT.
- RPE was recorded after 1,632 (97%) of the 1,675 attended treatment sessions (HIIT, 761/800 [95%]; MAT, 871/875 [~100%]). Compared with MAT, HIIT elicited significantly higher RPE overall.

Table S1. Treatment Intensity. Values are model estimates [95% CI]. Speed data are peak values within a bout. Heart rate data are presented as steady-state mean (excluding first 3 minutes) or peak values within a bout. Lactate, step counts and perceived exertion data are singular session values. All data are averaged across sessions and participants.

	HIIT (N=27)	MAT (N=28)	HIIT - MAT	p-value
Peak bout training speed, m/s				
All bouts	1.23 [1.07, 1.39]	0.76 [0.60, 0.91]	0.47 [0.25, 0.70]	0.0001
Overground bouts	1.20 [1.04, 1.36]	0.78 [0.63, 0.94]	0.41 [0.19, 0.64]	0.0005
Overground 1	1.21 [1.06, 1.37]	0.79 [0.64, 0.95]	0.42 [0.20, 0.64]	0.0004
Overground 2	1.18 [1.02, 1.34]	0.78 [0.62, 0.93]	0.41 [0.18, 0.63]	0.0006
Treadmill bout	1.30 [1.14, 1.46]	0.71 [0.55, 0.86]	0.59 [0.37, 0.82]	<0.0001
Treadmill - overground	0.10 [0.08, 0.12]	-0.08 [-0.09, -0.07]	0.18 [0.16, 0.20]	<0.0001
Mean change per session	0.01 [0.01, 0.01]	0.01 [0.01, 0.01]	0.00 [-0.00, 0.00]	0.4427
Block 1 (sessions 1-12)	1.14 [0.98, 1.29]	0.66 [0.51, 0.82]	0.47 [0.25, 0.69]	<0.0001
Block 2 (sessions 13-24)	1.25 [1.09, 1.41]	0.77 [0.62, 0.93]	0.48 [0.26, 0.70]	<0.0001
Block 3 (sessions 25-36)	1.31 [1.15, 1.47]	0.84 [0.68, 0.99]	0.47 [0.25, 0.70]	<0.0001
Peak bout training speed, % baseline fastest 10-m gait speed				
All bouts	161 [140, 182]	96 [75, 116]	65 [36, 95]	<0.0001
Overground bouts	150 [128, 171]	99 [78, 119]	51 [21, 81]	0.0011
Overground 1	153 [132, 175]	99 [79, 120]	54 [24, 84]	0.0006
Overground 2	146 [125, 167]	98 [77, 119]	48 [19, 78]	0.0020
Treadmill bout	184 [163, 205]	90 [69, 111]	94 [65, 124]	<0.0001
Treadmill - overground	35 [30, 39]	-9 [-11, -6]	43 [38, 48]	<0.0001
Mean change per session	1.5 [1.2, 1.7]	0.9 [0.7, 1.1]	0.6 [0.3, 0.8]	<0.0001
Block 1 (sessions 1-12)	142 [121, 164]	85 [64, 106]	57 [27, 87]	0.0003
Block 2 (sessions 13-24)	165 [144, 187]	96 [75, 117]	69 [39, 99]	<0.0001
Block 3 (sessions 25-36)	176 [155, 197]	106 [85, 127]	70 [40, 100]	<0.0001

Mean bout steady-state HR, %HRR$_{Tx\text{-}monitor}$ (relative to standing HR$_{rest}$ and most current HR$_{peak}$)				
All bouts	61 [57, 65]	46 [42, 49]	15 [10, 21]	<0.0001
Overground bouts	60 [56, 64]	44 [40, 48]	16 [10, 21]	<0.0001
Overground 1	53 [49, 57]	40 [36, 44]	13 [8, 19]	<0.0001
Overground 2	66 [62, 70]	48 [44, 52]	18 [13, 23]	<0.0001
Treadmill bout	64 [60, 68]	49 [45, 53]	15 [10, 20]	<0.0001
Treadmill - overground	4.2 [3.6, 4.8]	5.0 [4.4, 5.5]	-0.8 [-1.6, 0.0]	0.0639
Mean change per session	0.01 [-0.06, 0.07]	0.24 [0.19, 0.30]	-0.24 [-0.33, -0.15]	<0.0001
Block 1 (sessions 1-12)	62 [58, 66]	43 [39, 46]	19 [14, 25]	<0.0001
Block 2 (sessions 13-24)	61 [57, 64]	46 [42, 50]	14 [9, 20]	<0.0001
Block 3 (sessions 25-36)	61 [57, 65]	48 [44, 52]	13 [7, 18]	<0.0001
Peak bout HR, %HRR$_{Tx\text{-}monitor}$ (relative to standing HR$_{rest}$ and most current HR$_{peak}$)				
All bouts	80 [75, 86]	60 [55, 65]	21 [13, 28]	<0.0001
Overground bouts	77 [72, 83]	57 [52, 62]	20 [13, 28]	<0.0001
Overground 1	72 [66, 77]	52 [47, 58]	19 [12, 27]	<0.0001
Overground 2	83 [78, 88]	61 [56, 67]	22 [14, 29]	<0.0001
Treadmill bout	86 [81, 92]	66 [60, 71]	21 [13, 28]	<0.0001
Treadmill - overground	9.0 [8.1, 9.9]	8.6 [7.9, 9.4]	0.4 [-0.8, 1.6]	0.5050
Mean change per session	-0.24 [-0.33, -0.15]	0.10 [0.03, 0.17]	-0.34 [-0.45, -0.23]	<0.0001
Block 1 (sessions 1-12)	85 [79, 90]	59 [54, 64]	26 [18, 34]	<0.0001
Block 2 (sessions 13-24)	79 [73, 84]	60 [55, 65]	19 [11, 26]	<0.0001
Block 3 (sessions 25-36)	77 [72, 83]	61 [55, 66]	17 [9, 24]	0.0001
Mean bout steady-state HR, %HRR (relative to true HR$_{rest}$ and baseline HR$_{peak}$)				
All bouts	75 [70, 81]	59 [53, 64]	17 [10, 24]	<0.0001
Overground bouts	74 [69, 79]	57 [52, 62]	17 [10, 24]	<0.0001
Overground 1	67 [62, 73]	53 [48, 58]	14 [7, 22]	0.0002
Overground 2	81 [75, 86]	61 [56, 66]	20 [13, 27]	<0.0001
Treadmill bout	78 [73, 84]	62 [57, 67]	16 [9, 24]	<0.0001
Treadmill - overground	4.4 [3.8, 5.0]	5.3 [4.7, 5.9]	-0.8 [-1.7, 0.0]	0.0571
Mean change per session	0.59 [0.52, 0.66]	0.50 [0.45, 0.55]	0.09 [0.01, 0.18]	0.0369
Block 1 (sessions 1-12)	67 [62, 72]	52 [47, 57]	15 [8, 22]	0.0002
Block 2 (sessions 13-24)	77 [72, 82]	59 [54, 64]	18 [11, 26]	<0.0001
Block 3 (sessions 25-36)	82 [77, 87]	65 [59, 70]	17 [10, 25]	<0.0001
Peak bout HR, %HRR (relative to true HR$_{rest}$ and baseline HR$_{peak}$)				
All bouts	94 [87, 100]	71 [65, 78]	22 [13, 32]	<0.0001
Overground bouts	91 [84, 97]	68 [62, 75]	22 [13, 32]	<0.0001
Overground 1	85 [78, 92]	64 [58, 71]	21 [12, 30]	<0.0001
Overground 2	96 [90, 103]	73 [66, 79]	24 [14, 33]	<0.0001
Treadmill bout	100 [93, 106]	77 [71, 84]	23 [13, 32]	<0.0001
Treadmill - overground	9.1 [8.2, 9.9]	8.7 [7.9, 9.6]	0.3 [-0.9, 1.5]	0.5849
Mean change per session	0.50 [0.43, 0.58]	0.49 [0.43, 0.55]	0.02 [-0.08, 0.11]	0.7473
Block 1 (sessions 1-12)	87 [80, 93]	65 [59, 72]	22 [12, 31]	<0.0001
Block 2 (sessions 13-24)	95 [88, 102]	71 [65, 78]	23 [14, 33]	<0.0001
Block 3 (sessions 25-36)	99 [93, 106]	77 [71, 84]	22 [13, 32]	<0.0001
Blood lactate after treadmill bout, mmol/L				
After treadmill bout	3.1 [2.6, 3.6]	2.3 [1.9, 2.7]	0.8 [0.2, 1.5]	0.0141
Mean change per session	0.04 [-0.01, 0.10]	0.01 [-0.02, 0.03]	0.04 [-0.02, 0.10]	0.2285
Block 1 (sessions 1-12)	2.9 [2.3, 3.6]	2.1 [1.7, 2.6]	0.8 [0.0, 1.5]	0.0485
Block 2 (sessions 13-24)	3.0 [2.5, 3.5]	2.4 [1.9, 2.9]	0.6 [-0.1, 1.3]	0.0767
Block 3 (sessions 25-36)	3.4 [2.8, 3.9]	2.3 [1.9, 2.7]	1.1 [0.4, 1.8]	0.0014
Session step count				
During session	2,847 [2,529, 3,165]	3,532 [3,070, 3,993]	-685 [-1,244, -125]	0.0166
Mean change per session	13 [9, 16]	24 [21, 27]	-11 [-16, -7]	<0.0001
Block 1 (sessions 1-12)	2,656 [2,364, 2,948]	3,227 [2,737, 3,716]	-570 [-1,139, -2]	0.0494
Block 2 (sessions 13-24)	2,915 [2,596, 3,235]	3,576 [3,124, 4,028]	-661 [-1,213, -108]	0.0192
Block 3 (sessions 25-36)	2,970 [2,620, 3,319]	3,792 [3,344, 4,240]	-822 [-1,389, -255]	0.0046

Session rating of perceived exertion, 6-20				
After session	14.4 [13.8, 15.0]	13.4 [12.8, 14.1]	1.0 [0.1, 1.9]	0.0340
Mean change per session	0.03 [0.02, 0.04]	0.02 [0.01, 0.03]	0.01 [-0.01, 0.02]	0.3413
Block 1 (sessions 1-12)	14.0 [13.5, 14.6]	13.1 [12.5, 13.7]	0.9 [0.1, 1.7]	0.0255
Block 2 (sessions 13-24)	14.4 [13.8, 15.0]	13.7 [13.0, 14.3]	0.8 [-0.1, 1.6]	0.0787
Block 3 (sessions 25-36)	14.7 [14.0, 15.4]	13.5 [12.7, 14.3]	1.2 [0.2, 2.3]	0.0262

HR, heart rate; HRR, heart rate reserve; HR_{rest}, resting HR; HR_{peak}, highest HR achieved during exercise testing

Adverse events

There were no serious adverse events related to study procedures and no significant between-group differences in any adverse event categories (Table S2). Four participants experienced serious adverse events during study participation that were all determined to be unrelated to study procedures, including: a seizure leading to temporary hospitalization (n=1, MAT group), a fall with hip fracture (n=1, MAT group), an episode of delirium leading to temporary hospitalization (n=1, MAT group) and a recurrent stroke (n=1, HIIT group).

Table S2. Adverse Events (AEs)
Values are N participants with AE (total N AEs).

	HIIT (N=27) 800 sessions	**MAT** (N=28) 875 sessions	**HIIT/MAT odds ratio [95% CI]**[a]
Any post-randomization AE	25 (109)	23 (84)	3.2 [0.5, 20.5]
-Grade 1 (mild)	22 (80)	22 (62)	1.2 [0.3, 5.2]
-Grade 2 (moderate)	14 (28)	9 (19)	2.7 [0.8, 9.1]
-Grade 3 (severe)	1 (1)	3 (3)	0.3 [0.0, 3.4]
-Grade 4-5 (life threat-death)	0 (0)	0 (0)	1 [NA]
-Pain/Soreness	12 (25)	11 (32)	1.3 [0.4, 4.0]
-Fatigue	4 (4)	5 (6)	0.8 [0.2, 3.6]
-Lightheadedness	6 (17)	7 (15)	0.9 [0.2, 3.3]
-Fall	9 (36)	11 (14)	0.7 [0.2, 2.3]
Definitely treatment-related	13 (39)	10 (29)	1.7 [0.6, 5.2]
-Grade 1 (mild)	11 (26)	9 (24)	1.5 [0.5, 4.7]
-Grade 2 (moderate)	7 (13)	4 (5)	2.1 [0.5, 8.4]
-Grade 3-5 (severe-death)	0 (0)	0 (0)	1 [NA]
-Pain/Soreness	8 (15)	7 (18)	1.3 [0.4, 4.4]
-Fatigue	2 (2)	2 (2)	1.1 [0.1, 8.7]
-Lightheadedness	3 (10)	4 (8)	0.8 [0.1, 4.0]
-Fall	3 (4)	0 (0)	8.2 [0.4, 167][b]
Possibly/not treatment-related	22 (70)	19 (55)	2.3 [0.6, 8.5]
-Grade 1 (mild)	19 (54)	17 (38)	1.6 [0.5, 5.2]
-Grade 2 (moderate)	10 (15)	7 (14)	2.4 [0.6, 9.7]
-Grade 3 (severe)	1 (1)	3 (3)	0.3 [0.0, 3.4]
-Grade 4-5 (life threat-death)	0 (0)	0 (0)	1 [NA]
-Pain/Soreness	7 (10)	8 (14)	0.9 [0.3, 3.2]
-Fatigue	2 (2)	3 (4)	0.7 [0.1, 5.9]
-Lightheadedness	3 (7)	5 (7)	0.6 [0.1, 2.9]
-Fall	9 (32)	11 (14)	0.7 [0.2, 2.3]

[a]*Relative odds of a participant having an AE, adjusted for site and baseline walking limitation severity.* [b]*When AEs were only present in one group, continuity correction added 0.5 AEs to each group to permit (unadjusted) ratio calculation.*

Other exercise test measures to facilitate interpretation and exploratory outcomes

Other exercise capacity outcomes showed significant increases in both groups at various time points (Table S3), with some measures increasing significantly more for HIIT versus MAT at 4WK (GXT duration, peak GXT speed and peak GXT HR) or 12WK (peak GXT HR and HR_{peak}). The verification test increased HR_{peak} by 2.7 [1.0, 4.4] bpm at baseline and 2.0 [1.3, 2.7] bpm across all time points. Baseline HR_{peak} was 87.0 ± 17.2% age-predicted maximal HR (unadjusted mean ± SD). Peak respiratory exchange ratio and rating of perceived exertion during the baseline GXT were 1.01 ± 0.11 and 15.2 ± 2.6, respectively, and did not significantly change over time in either group. Meanwhile, there were significant improvements in both groups for ventilatory threshold treadmill speed, metabolic cost of gait in the last 3 minutes of the GXT, and functional ambulation category score, with no significant differences between groups. Only the HIIT group reported significant improvement in overall quality of life (lower EuroQoL misery score) and performance of usual activities (lower EuroQoL score for that item), and this was only at 4WK, with no significant between-group differences after FDR correction.

Table S3. Other exercise test measures and exploratory outcomes

	HIIT (N=27)	MAT (N=28)	HIIT - MAT	p-value
Graded Exercise Test (GXT) Duration, minutes				
Baseline	9.6 (4.2)	8.9 (4.2)	0.7 [-1.4, 2.8]	0.5212
4-Week Change	1.8 [1.1, 2.4]	0.6 [-0.1, 1.2]	1.2 [0.3, 2.1]	0.0078[a]
8-Week Change	2.5 [1.5, 3.4]	1.6 [0.6, 2.5]	0.9 [-0.3, 2.1]	0.1537
12-Week Change	3.3 [2.2, 4.4]	2.2 [1.1, 3.4]	1.1 [-0.4, 2.6]	0.1350
Peak Treadmill Speed During GXT, m/s				
Baseline	0.69 (0.36)	0.64 (0.37)	0.05 [-0.14, 0.23]	0.6081
4-Week Change	0.15 [0.09, 0.22]	0.04 [-0.03, 0.11]	0.11 [0.03, 0.20]	0.0093[a]
8-Week Change	0.17 [0.09, 0.26]	0.13 [0.04, 0.22]	0.04 [-0.07, 0.16]	0.4391
12-Week Change	0.24 [0.14, 0.33]	0.20 [0.10, 0.30]	0.04 [-0.09, 0.16]	0.5570
Peak Oxygen Consumption Rate During GXT ($VO_{2\text{-peak}}$), mL/kg/min				
Baseline	13.4 (4.4)	13.0 (4.4)	0.4 [-1.8, 2.6]	0.7160
4-Week Change	1.5 [0.5, 2.5]	0.2 [-0.8, 1.2]	1.3 [0.0, 2.6]	0.0478
8-Week Change	2.1 [0.8, 3.5]	1.5 [0.0, 2.9]	0.7 [-1.2, 2.5]	0.4778
12-Week Change	2.5 [0.8, 4.2]	1.5 [-0.3, 3.3]	1.1 [-1.2, 3.4]	0.3562
Peak Heart Rate During GXT, bpm				
Baseline	123.4 (21.1)	128.2 (21.3)	-4.9 [-15.5, 5.8]	0.3663
4-Week Change	5.4 [0.9, 9.9]	-3.7 [-8.2, 0.7]	9.1 [3.3, 15.0]	0.0030[a]
8-Week Change	4.2 [-1.5, 9.9]	-2.2 [-8.2, 3.9]	6.4 [-1.3, 14.0]	0.0994
12-Week Change	12.6 [6.5, 18.6]	-1.0 [-7.2, 5.2]	13.6 [5.6, 21.6]	0.0013[a]
Highest Heart Rate During GXT or Verification Test (HR_{peak}), bpm				
Baseline	125.0 (22.1)	131.4 (22.3)	-6.4 [-17.6, 4.8]	0.2562
4-Week Change	5.7 [0.2, 11.1]	-3.7 [-9.2, 1.7]	9.4 [2.2, 16.5]	0.0111
8-Week Change	4.1 [-1.7, 10.0]	-3.4 [-9.6, 2.8]	7.5 [-0.3, 15.4]	0.0605
12-Week Change	11.2 [4.7, 17.8]	-2.6 [-9.3, 4.0]	13.9 [5.2, 22.5]	0.0023[a]
HR_{peak}, % Age-Predicted Maximal HR				
Baseline	83.3 (18.4)	90.0 (18.5)	-6.7 [-16.0, 2.6]	0.1528
4-Week Change	3.8 [0.2, 7.4]	-2.4 [-6.0, 1.2]	6.2 [1.4, 11.0]	0.0119
8-Week Change	3.4 [-0.7, 7.4]	-2.1 [-6.4, 2.2]	5.4 [0.0, 10.9]	0.0497
12-Week Change	7.8 [3.1, 12.6]	-2.2 [-7.0, 2.7]	10.0 [3.7, 16.3]	0.0026[a]
Peak Respiratory Exchange Ratio During GXT (VCO_2/VO_2)				
Baseline	1.01 (0.12)	1.02 (0.12)	-0.01 [-0.07, 0.05]	0.7949
4-Week Change	-0.01 [-0.04, 0.02]	-0.03 [-0.06, 0.00]	0.01 [-0.03, 0.05]	0.5181
8-Week Change	0.03 [-0.02, 0.07]	-0.02 [-0.06, 0.03]	0.04 [-0.01, 0.10]	0.1432
12-Week Change	0.00 [-0.04, 0.04]	-0.01 [-0.05, 0.03]	0.00 [-0.05, 0.06]	0.8689
Rating of Perceived Exertion (RPE) from GXT, 6-20				
Baseline	14.2 (2.5)	15.1 (2.5)	-0.9 [-2.1, 0.4]	0.1655
4-Week Change	0.4 [-1.0, 1.8]	-0.5 [-1.9, 0.8]	1.0 [-0.8, 2.7]	0.2796
8-Week Change	0.7 [-0.5, 1.9]	0.1 [-1.1, 1.3]	0.6 [-1.0, 2.1]	0.4559
12-Week Change	0.9 [-0.2, 2.0]	0.6 [-0.5, 1.6]	0.3 [-1.0, 1.7]	0.6185

Ventilatory Threshold Treadmill Speed, m/s				
Baseline	0.60 (0.32)	0.53 (0.33)	0.07 [-0.09, 0.23]	0.3972
4-Week Change	0.13 [0.07, 0.18]	0.06 [0.00, 0.11]	0.07 [-0.01, 0.14]	0.0691
8-Week Change	0.16 [0.09, 0.23]	0.12 [0.05, 0.19]	0.04 [-0.05, 0.13]	0.3615
12-Week Change	0.21 [0.13, 0.28]	0.19 [0.11, 0.27]	0.02 [-0.08, 0.12]	0.7148
Metabolic Cost of Treadmill Gait During Last 3 Minutes of GXT, mLO_2/kg/m				
Baseline	0.44 (0.19)	0.47 (0.19)	-0.03 [-0.12, 0.07]	0.5372
4-Week Change	-0.08 [-0.13, -0.03]	-0.05 [-0.10, 0.00]	-0.03 [-0.10, 0.03]	0.3293
8-Week Change	-0.10 [-0.14, -0.05]	-0.08 [-0.13, -0.03]	-0.01 [-0.08, 0.05]	0.6751
12-Week Change	-0.12 [-0.17, -0.06]	-0.11 [-0.17, -0.06]	0.00 [-0.07, 0.06]	0.9025
Metabolic Cost of Treadmill Gait During Last 3 Minutes of *Shortest* GXT, mLO_2/kg/m				
Baseline	0.45 (0.18)	0.48 (0.18)	-0.03 [-0.12, 0.06]	0.5095
4-Week Change	-0.01 [-0.04, 0.02]	-0.01 [-0.04, 0.02]	0.00 [-0.04, 0.04]	0.9400
8-Week Change	-0.02 [-0.06, 0.02]	-0.04 [-0.08, 0.00]	0.02 [-0.03, 0.08]	0.4346
12-Week Change	-0.01 [-0.06, 0.03]	-0.04 [-0.08, 0.01]	0.02 [-0.04, 0.08]	0.4298
Gait Stability: Functional Ambulation Category, 0-4				
Baseline	3.16 (0.58)	3.19 (0.58)	-0.03 [-0.32, 0.26]	0.8305
4-Week Change	-0.02 [-0.24, 0.20]	0.17 [-0.05, 0.39]	-0.19 [-0.48, 0.10]	0.1950
8-Week Change	0.22 [-0.03, 0.46]	0.31 [0.06, 0.56]	-0.09 [-0.42, 0.24]	0.5727
12-Week Change	0.26 [0.02, 0.50]	0.27 [0.03, 0.51]	-0.01 [-0.33, 0.31]	0.9475
Activities-Specific Balance Confidence (ABC) Scale Score, 0-100				
Baseline	59.6 (18.9)	63.0 (19.1)	-3.4 [-12.9, 6.2]	0.4815
4-Week Change	-2.7 [-7.2, 1.7]	-4.6 [-9.1, -0.1]	1.9 [-4.0, 7.7]	0.5266
8-Week Change	1.6 [-3.1, 6.4]	-1.8 [-6.7, 3.1]	3.5 [-2.8, 9.7]	0.2757
12-Week Change	2.5 [-2.7, 7.7]	0.6 [-4.7, 5.9]	1.9 [-5.0, 8.8]	0.5758
EuroQOL-5D-5L Misery Score, 5-25 (higher score = worse self-reported health problems)				
Baseline	11.0 (2.7)	10.7 (2.7)	0.3 [-1.1, 1.7]	0.6669
4-Week Change	-0.9 [-1.8, -0.1]	-0.2 [-1.1, 0.7]	-0.7 [-1.9, 0.4]	0.2158
8-Week Change	-0.6 [-1.5, 0.3]	-0.4 [-1.3, 0.5]	-0.2 [-1.3, 1.0]	0.7859
12-Week Change	-0.1 [-1.0, 0.8]	-0.6 [-1.5, 0.3]	0.5 [-0.7, 1.7]	0.3777
EuroQOL-5D-5L Mobility Score, 1-5 (higher score = worse self-reported problems)				
Baseline	2.56 (0.86)	2.51 (0.87)	0.05 [-0.38, 0.49]	0.8081
4-Week Change	-0.17 [-0.58, 0.24]	-0.15 [-0.56, 0.26]	-0.01 [-0.55, 0.52]	0.9573
8-Week Change	-0.17 [-0.60, 0.27]	-0.13 [-0.58, 0.32]	-0.04 [-0.62, 0.54]	0.8921
12-Week Change	-0.23 [-0.64, 0.18]	-0.24 [-0.65, 0.17]	0.01 [-0.53, 0.54]	0.9824
EuroQOL-5D-5L Usual Activities Score, 1-5 (higher score = worse self-reported problems)				
Baseline	2.53 (0.88)	2.23 (0.89)	0.30 [-0.15, 0.74]	0.1822
4-Week Change	-0.41 [-0.71, -0.11]	0.02 [-0.28, 0.31]	-0.43 [-0.82, -0.04]	0.0320
8-Week Change	-0.25 [-0.63, 0.14]	0.11 [-0.29, 0.52]	-0.36 [-0.88, 0.15]	0.1615
12-Week Change	-0.16 [-0.57, 0.24]	-0.07 [-0.48, 0.34]	-0.09 [-0.63, 0.44]	0.7262

Baseline data presented as (covariate-adjusted) mean (SD)
Change and difference data presented as mean difference [95% CI]
[a]$p<0.05$ after false discovery rate-correction across timepoints & all measures

Primary outcome changes stratified by baseline walking limitation severity

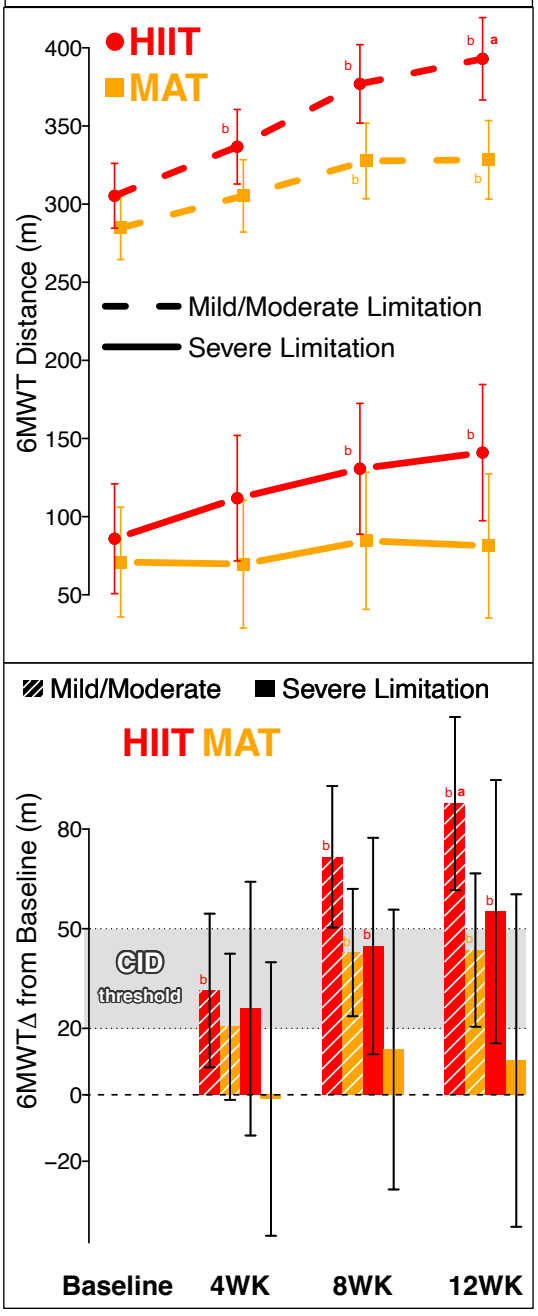
Among the 41 participants (20 HIIT, 21 MAT) with mild/moderate baseline walking limitations (self-selected speed 0.4-1.0 m/s), 6MWT distance significantly increased in both the HIIT and MAT groups, with significantly greater improvement in HIIT at 12WK (Figure S3 & Table S4). Among the 14 participants (7 HIIT, 7 MAT) with severe baseline walking limitations (self-selected speed <0.4 m/s), 6MWT distance significantly increased in the HIIT group only, but with no significant difference between groups.

Table S4. Primary outcome (6-Minute Walk Distance) changes stratified by baseline walking limitation severity

	HIIT (N=20)	MAT (N=21)	HIIT - MAT	p-value
Mild/Moderate Baseline Limitation (gait speed 0.4-1.0 m/s; N=41)				
Baseline	305 (93)	285 (93)	21 [-38, 79]	0.4816
4-Week Change	31 [8, 55]	20 [-2, 42]	11 [-21, 43]	0.4960
8-Week Change	72 [50, 93]	43 [24, 62]	29 [0, 57]	0.0490
12-Week Change	88 [62, 114]	44 [20, 67]	44 [9, 79]	0.0142[a]
Severe Baseline Limitation (gait speed <0.4 m/s; N=14)				
	HIIT (N=7)	**MAT (N=7)**	**HIIT - MAT**	**p-value**
Baseline	86 (93)	71 (93)	15 [-85, 115]	0.7656
4-Week Change	26 [-12, 64]	-1 [-42, 40]	27 [-29, 83]	0.3348
8-Week Change	45 [12, 77]	14 [-29, 56]	31 [-22, 85]	0.2464
12-Week Change	55 [15, 95]	10 [-40, 60]	45 [-19, 109]	0.1656

Baseline data presented as (covariate-adjusted) mean (SD)
Change and difference data presented as mean difference [95% CI]
[a]p<0.05 after false discovery rate-correction across time points

Figure S3. Primary analysis stratified by baseline walking limitation severity. Values are model estimates and error bars are 95% CI. Mild/moderate limitation, gait speed 0.4-1.0 m/s; severe limitation, <0.4 m/s. [a]pFDR<0.05 between-group difference in change (Δ) from baseline. [b]p<0.05 within-group Δ from baseline. FDR, false discovery rate corrected across time points; 6MWT, 6-minute walk test; CID, clinically important difference; WK, week



Sensitivity of primary results to missing data assumptions

When assuming that participants who dropped out due to an adverse event had no mean improvement (rather than the common "missing at random" assumption), both groups still showed significant improvements in 6MWT distance and the direction of the between-group differences remained unchanged, but the between-group effect sizes were smaller and no longer statistically significant (Table S5).

Table S5. Estimated 6-minute walk test changes when assuming no mean improvement in cases of adverse-event-related dropout

	HIIT (N=27)	MAT (N=28)	HIIT - MAT	p-value
Walking Capacity: 6-Minute Walk Distance, m *(primary outcome)*				
Baseline	196 (98)	177 (99)	19 [-31, 69]	0.4438
4-Week Change	26 [7, 46]	13 [-7, 32]	14 [-12, 40]	0.2992
8-Week Change	47 [28, 66]	28 [9, 47]	19 [-6, 44]	0.1325
12-Week Change	52 [30, 73]	28 [6, 49]	24 [-4, 53]	0.0985

Baseline data presented as (covariate-adjusted) mean (SD)
Change and difference data presented as mean difference [95% CI]

Supplementary Discussion

Additional discussion of primary results

Within the MAT group, the mean 6MWT gain after 12 weeks of training was comparable to many similar previous studies in chronic stroke,[11, 62-67] and exceeded the minimally important difference of +20 meters.[29] This confirms that MAT was successfully implemented in this study and has some value for improving walking capacity in this population. Surprisingly, our results suggest that most outcomes reached an apparent plateau after just 8 weeks of MAT, which may not support the longer 3-6 month training durations used in most previous stroke trials focused on MAT. However, there are prior indications that outcomes may continue to slowly improve between 3-6 months of MAT.[11, 68] In addition, it is plausible that longer training durations could result in more sustained gains in walking capacity or greater long-term improvements in cardiovascular-metabolic health, but this has not been thoroughly tested.

Discussion of supplementary results

Among the exploratory outcomes, HIIT elicited significantly greater improvements than MAT in exercise capacity (GXT duration, peak GXT speed and peak HR) at various time points after adjustment for multiple comparisons (Table S3).

Our prespecified subgroup analysis (Figure S3 & Table S4) suggested that the vigorous training intensity of HIIT may have been particularly important for participants with severe baseline walking limitations (self-selected gait speed <0.4 m/s[2, 58]). For these individuals, MAT elicited mean 6MWT gains less than the minimally important difference of +20 meters[29] (+10 [-40-60] meters). Previous research also found minimal response to MAT among stroke survivors with severe walking limitations,[6, 58] leading to prior suggestions that these patients may not be good candidates for gait training.[58] However, in the current study, participants with severe walking limitations increased mean 6MWT distance by +55 [15-95] meters with HIIT, which is above the +50-meter threshold for a large clinically-meaningful gain.[29] This study was not powered to detect significant differences between HIIT and MAT within the severe limitation subgroup (N=14, p=0.17) or to assess within-group changes with sufficient precision to rule meaningful benefits in or out with confidence, so these findings should be taken as preliminary. However, HIIT did show significantly greater improvement than MAT in the larger mild/moderate limitation subgroup (+88 [62-114] vs +44 [20-67], N=41, p=0.01), suggesting that vigorous training intensity could be optimal regardless of baseline limitation.

As expected, the symptom-limited GXT was primarily terminated by non-cardiorespiratory factors on average (respiratory exchange ratio < 1.05, RPE < 17, HR $\lesssim$ 90% age-predicted max), so GXT duration and peak GXT measures cannot be interpreted as specific measures of cardiopulmonary (aerobic) capacity, but can still be interpreted as measures of *exercise capacity*, which can be influenced by neuromotor function, cardiopulmonary capacity and other factors (e.g. motivation).

Importantly, significant improvements in the metabolic cost of gait and gait stability for both groups suggest that compensatory movements and compromised stability were not the primary mechanisms by which individuals achieved gains in walking capacity with training. However, it did appear that gains in efficiency were primarily related to gains in gait speed, since efficiency at matched speeds did not significantly change.

Additional discussion of strengths and limitations

Another major strength of this trial from this supplementary appendix is that we successfully implemented several novel technological methods to enhance protocol fidelity, including automated real-time prompts, calculations and feedback from the direct electronic data entry system during all visits, plus automated bout timing, recording & intensity guidance from a fully wireless HR monitoring system during treatment visits.

Similar to the PROMIS-Fatigue scale, the lack of follow up testing also makes it more challenging to interpret the results of the exploratory questionnaires (ABC scale and EuroQOL-5D-5L) reported in this appendix since they were asking participants to report about time periods in which they were engaged in the study treatment at 4WK, 8WK and 12WK, but asking about a time period when participants were largely just participating in normal daily activities at baseline. We attempted to mitigate this issue by specifically asking participants to answer questions based on their usual activities outside of the study, but are skeptical that this strategy was completely successful. Thus, a larger trial with post-treatment follow-up testing is needed to confirm and expand on these results.

eReferences

1. Macko RF, Ivey FM, Forrester LW. Task-oriented aerobic exercise in chronic hemiparetic stroke: training protocols and treatment effects. *Topics in Stroke Rehabilitation*. 2005;12(1):45-57.

2. Boyne P, Doren S, Scholl V, et al. Preliminary Outcomes of Combined Treadmill and Overground High-Intensity Interval Training in Ambulatory Chronic Stroke. *Frontiers in Neurology*. 2022;13:812875.

3. Boyne P, Scholl V, Doren S, et al. Locomotor training intensity after stroke: Effects of interval type and mode. *Top Stroke Rehabil*. 2020;27(7):483-493. doi: 10.1080/10749357.2020.1728953 [doi].

4. American College of Sports Medicine, (ACSM). *ACSM's guidelines for exercise testing and prescription.* 9th ed. Philadephia, PA: Lippincott Williams & Wilkins; 2014.

5. Boyne P, Dunning K, Carl D, Gerson M, Khoury J, Kissela B. Within-session responses to high-intensity interval training in chronic stroke. *Med Sci Sports Exerc*. 2015;47(3):476-484. doi: 10.1249/MSS.0000000000000427 [doi].

6. Boyne P, Dunning K, Carl D, et al. High-Intensity Interval Training and Moderate-Intensity Continuous Training in Ambulatory Chronic Stroke: A Feasibility Study. *Phys Ther*. 2016;96(10):1533-1544. doi: ptj.20150277 [pii].

7. Billinger SA, Arena R, Bernhardt J, et al. Physical Activity and Exercise Recommendations for Stroke Survivors: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2014;45(8):2532-53. doi: STR.0000000000000022 [pii].

8. Winstein CJ, Stein J, Arena R, et al. Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2016;47(6):e98-e169. doi: 10.1161/STR.0000000000000098 [doi].

9. Macko RF, Haeuber E, Shaughnessy M, et al. Microprocessor-based ambulatory activity monitoring in stroke patients. *Med Sci Sports Exerc*. 2002;34(3):394-399.

10. Miller A, Reisman DS, Billinger SA, et al. Moderate-Intensity Exercise Versus High-Intensity Interval Training to Recover Walking Post-Stroke: Protocol for a Randomized Controlled Trial. *Trials*. 2021;22(1):457. doi: 10.1186/s13063-021-05419-x.

11. Ada L, Dean CM, Lindley R. Randomized trial of treadmill training to improve walking in community-dwelling people after stroke: the AMBULATE trial. *Int J Stroke*. 2013;8(6):436-444. doi: 10.1111/j.1747-4949.2012.00934.x [doi].

12. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166(1):111-117.

13. Boyne P, Dunning K, Carl D, Gerson M, Khoury J, Kissela B. High-intensity interval training in stroke rehabilitation. *Top Stroke Rehabil*. 2013;20(4):317-330. doi: 10.1310/tsr2004-317; 10.1310/tsr2004-317.

14. Mudge S, Stott NS. Timed walking tests correlate with daily step activity in persons with stroke. *Arch Phys Med Rehabil*. 2009;90(2):296-301. doi: 10.1016/j.apmr.2008.07.025 [doi].

15. Fulk GD, Reynolds C, Mondal S, Deutsch JE. Predicting home and community walking activity in people with stroke. *Arch Phys Med Rehabil*. 2010;91(10):1582-1586. doi: 10.1016/j.apmr.2010.07.005.

16. Donovan K, Lord SE, McNaughton HK, Weatherall M. Mobility beyond the clinic: the effect of environment on gait and its measurement in community-ambulant stroke survivors. *Clin Rehabil*. 2008;22(6):556-563. doi: 10.1177/0269215507085378 [doi].

17. Miller AE, Russell E, Reisman DS, Kim HE, Dinh V. A machine learning approach to identifying important features for achieving step thresholds in individuals with chronic stroke. *PLoS One*. 2022;17(6):e0270105. doi: 10.1371/journal.pone.0270105 [doi].

18. Muren MA, Hutler M, Hooper J. Functional capacity and health-related quality of life in individuals post stroke. *Top Stroke Rehabil*. 2008;15(1):51-58. doi: 10.1310/tsr1501-51.

19. Rich S, MD. The 6-Minute Walk Test as a Primary Endpoint in Clinical Trials for Pulmonary Hypertension. *J Am Coll Cardiol*. 2012;60(13):1202-1203. doi: 10.1016/j.jacc.2012.03.080.

20. Muenzer J, Wraith JE, Beck M, et al. A phase II/III clinical study of enzyme replacement therapy with idursulfase in mucopolysaccharidosis II (Hunter syndrome). *Genetics in medicine*. 2006;8(8):465-473. doi: 10.1097/01.gim.0000232477.37660.fb.

21. Wraith JE, Clarke LA, Beck M, et al. Enzyme replacement therapy for mucopolysaccharidosis I: a randomized, double-blinded, placebo-controlled, multinational study of recombinant human α-L-iduronidase (laronidase). *J Pediatr*. 2004;144(5):581-588. doi: 10.1016/j.jpeds.2004.01.046.

22. van der Ploeg, Ans T., Clemens PR, Corzo D, et al. A Randomized Study of Alglucosidase Alfa in Late-Onset Pompe's Disease. *N Engl J Med*. 2010;362(15):1396-1406. doi: 10.1056/NEJMoa0909859.

23. Raval AN, Cook TD, Duckers HJ, et al. The CardiAMP Heart Failure trial: A randomized controlled pivotal trial of high-dose autologous bone marrow mononuclear cells using the CardiAMP cell therapy system in patients with post–myocardial infarction heart failure: Trial rationale and study design. *The American heart journal*. 2018;201:141-148. doi: 10.1016/j.ahj.2018.03.016.

24. Fulk GD, Echternach JL, Nof L, O'Sullivan S. Clinometric properties of the six-minute walk test in individuals undergoing rehabilitation poststroke. *Physiother Theory Pract*. 2008;24(3):195-204. doi: 10.1080/09593980701588284.

25. Eng JJ, Dawson AS, Chu KS. Submaximal exercise in persons with stroke: test-retest reliability and concurrent validity with maximal oxygen consumption. *Arch Phys Med Rehabil*. 2004;85(1):113-118.

26. Wevers LE, Kwakkel G, van de Port, I G. Is outdoor use of the six-minute walk test with a global positioning system in stroke patients' own neighbourhoods reproducible and valid?. *J Rehabil Med*. 2011;43(11):1027-1031. doi: 10.2340/16501977-0881 [doi].

27. Flansbjer U, Holmbäck AM, Downham D, Patten C, Lexell J. Reliability of Gait Performance Tests in Men and Women with Hemiparesis After Stroke. *J Rehabil Med*. 2005;37(2):75-82.

28. Kosak M, Smith T. Comparison of the 2-, 6-, and 12-minute walk tests in patients with stroke. *J Rehabil Res Dev*. 2005;42(1):103-107.

29. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54(5):743-749. doi: JGS701 [pii].

30. Tilson JK, Sullivan KJ, Cen SY, et al. Meaningful gait speed improvement during the first 60 days poststroke: minimal clinically important difference. *Phys Ther*. 2010;90(2):196-208. doi: 10.2522/ptj.20090079.

31. Cella D, Lai J, Jensen SE, et al. PROMIS Fatigue Item Bank had Clinical Validity across Diverse Chronic Conditions. *J Clin Epidemiol*. 2016;73:128-134. doi: 10.1016/j.jclinepi.2015.08.037.

32. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63(11):1179-1194. doi: 10.1016/j.jclinepi.2010.04.011.

33. Boyne P, Reisman D, Brian M, et al. Ventilatory threshold may be a more specific measure of aerobic capacity than peak oxygen consumption rate in persons with stroke. *Top Stroke Rehabil*. 2017;24(2):149-157. doi: 10.1080/10749357.2016.1209831 [doi].

34. Marzolini S, Oh P, McIlroy W, Brooks D. The feasibility of cardiopulmonary exercise testing for prescribing exercise to people after stroke. *Stroke*. 2012;43(4):1075-1081. doi: 10.1161/STROKEAHA.111.635128; 10.1161/STROKEAHA.111.635128.

35. American Thoracic Society, American College of Chest Physicians. ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167(2):211-277. doi: 10.1164/rccm.167.2.211.

36. Mezzani A, Hamm LF, Jones AM, et al. Aerobic exercise intensity assessment and prescription in cardiac rehabilitation: a joint position statement of the European Association for Cardiovascular Prevention and Rehabilitation, the American Association of Cardiovascular and Pulmonary Rehabilitation, and the Canadian Association of Cardiac Rehabilitation. *J Cardiopulm Rehabil Prev*. 2012;32(6):327-350. doi: 10.1097/HCR.0b013e3182757050 [doi].

37. Marzolini S, Oh P, McIlroy W, Brooks D. The Effects of an Aerobic and Resistance Exercise Training Program on Cognition Following Stroke. *Neurorehabil Neural Repair*. 2013;27(5):392-402. doi: 10.1177/1545968312465192.

38. Holleran CL, Rodriguez KS, Echauz A, Leech KA, Hornby TG. Potential contributions of training intensity on locomotor performance in individuals with chronic stroke. *J Neurol Phys Ther*. 2015;39(2):95-102. doi: 10.1097/NPT.0000000000000077 [doi].

39. Kim KJ, Rivas E, Prejean B, Frisco D, Young M, Downs M. Novel Computerized Method for Automated Determination of Ventilatory Threshold and Respiratory Compensation Point. *Frontiers in physiology*. 2021;12:782167. doi: 10.3389/fphys.2021.782167.

40. Ekkekakis P, Lind E, Hall EE, Petruzzello SJ. Do regression-based computer algorithms for determining the ventilatory threshold agree?. *J Sports Sci*. 2008;26(9):967-976. doi: 10.1080/02640410801910269.

41. Gaskill SE, Ruby BC, Walker AJ, Sanchez OA, Serfass RC, Leon AS. Validity and reliability of combining three methods to determine ventilatory threshold. *Med Sci Sports Exerc*. 2001;33(11):1841-1848. doi: 10.1097/00005768-200111000-00007.

42. R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*. 2004.

43. Lindeløv JK. mcp: An R Package for Regression With Multiple Change Points. *OSF Preprints [preprint]*. 2020:https://doi-org.uc.idm.oclc.org/10.31219/osf.io/fzqxv.

44. Brawner CA, Ehrman JK, Schairer JR, Cao JJ, Keteyian SJ. Predicting maximum heart rate among patients with coronary heart disease receiving beta-adrenergic blockade therapy. *Am Heart J*. 2004;148(5):910-914. doi: 10.1016/j.ahj.2004.04.035.

45. Gellish RL, Goslin BR, Olson RE, McDonald A, Russi GD, Moudgil VK. Longitudinal modeling of the relationship between age and maximal heart rate. *Med Sci Sports Exerc*. 2007;39(5):822-829. doi: 10.1097/mss.0b013e31803349c6.

46. Tyrell CM, Roos MA, Rudolph KS, Reisman DS. Influence of systematic increases in treadmill walking speed on gait kinematics after stroke. *Phys Ther*. 2011;91(3):392-403. doi: 10.2522/ptj.20090425.

47. Leddy AL, Connolly M, Holleran CL, et al. Alterations in Aerobic Exercise Performance and Gait Economy Following High-Intensity Dynamic Stepping Training in Persons With Subacute Stroke. *J Neurol Phys Ther*. 2016;40(4):239-248. doi: 10.1097/NPT.0000000000000147 [doi].

48. Mehrholz J, Wagner K, Rutte K, Meissner D, Pohl M. Predictive validity and responsiveness of the functional ambulation category in hemiparetic patients after stroke. *Arch Phys Med Rehabil*. 2007;88(10):1314-1319. doi: 10.1016/j.apmr.2007.06.764.

49. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *J Gerontol A Biol Sci Med Sci*. 1995;50A(1):28.

50. Botner EM, Miller WC, Eng JJ. Measurement properties of the Activities-specific Balance Confidence Scale among individuals with stroke. *Disabil Rehabil*. 2005;27(4):156-163. doi: 10.1080/09638280400008982.

51. Golicki D, Niewada M, Buczek J, et al. Validity of EQ-5D-5L in stroke. *Quality of life research*. 2014;24(4):845-850. doi: 10.1007/s11136-014-0834-1.

52. Pickard AS, PhD, Law, Ernest H.,PharmD, PhD, Jiang R, PharmD, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value in health*. 2019;22(8):931-941. doi: 10.1016/j.jval.2019.02.009.

53. Bohannon RW, Williams Andrews A. Normal walking speed: a descriptive meta-analysis. *Physiotherapy*. 2011;97(3):182-189. doi: 10.1016/j.physio.2010.12.004.

54. Enright PL, Sherrill DL. Reference equations for the six-minute walk in healthy adults. *Am J Respir Crit Care Med*. 1998;158(5 Pt 1):1384-1387.

55. Guazzi M, Adams V, Conraads V, et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation*. 2012;126(18):2261-2274. doi: 10.1161/CIR.0b013e31826fb946; 10.1161/CIR.0b013e31826fb946.

56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B*. 1995;57(1):289-300.

57. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis.* 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2011.

58. Dean CM, Ada L, Lindley RI. Treadmill training provides greater benefit to the subgroup of community-dwelling people after stroke who walk faster than 0.4m/s: a randomised trial. *J Physiother*. 2014;60(2):97-101. doi: 10.1016/j.jphys.2014.03.004 [doi].

59. Higgins J, Savović J, Page MJ, Elbers RG, Sterne J. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins J, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3.* www.training.cochrane.org/handbook: Cochrane; 2022.

60. Duncan PW, Sullivan KJ, Behrman AL, et al. Protocol for the Locomotor Experience Applied Post-stroke (LEAPS) trial: a randomized controlled trial. *BMC Neurol*. 2007;7:39. doi: 10.1186/1471-2377-7-39.

61. Guo Y, Logan HL, Glueck DH, Muller KE. Selecting a sample size for studies with repeated measures. *BMC Med Res Methodol*. 2013;13:100-100. doi: 10.1186/1471-2288-13-100 [doi].

62. Ivey FM, Ryan AS, Hafer-Macko CE, Macko RF. Improved cerebral vasomotor reactivity after exercise training in hemiparetic stroke survivors. *Stroke*. 2011;42(7):1994-2000. doi: 10.1161/STROKEAHA.110.607879.

63. Luft AR, Macko RF, Forrester LW, et al. Treadmill Exercise Activates Subcortical Neural Networks and Improves Walking After Stroke. A Randomized Controlled Trial. *Stroke*. 2008;39(12):3341-50. doi: 10.1161/STROKEAHA.108.527531.

64. Olawale OA, Jaja SI, Anigbogu CN, Appiah-Kubi KO, Jones-Okai D. Exercise training improves walking function in an African group of stroke survivors: a randomized controlled trial. *Clin Rehabil*. 2011;25(5):442-450. doi: 10.1177/0269215510389199.

65. Lee J, Moon H, Lee S, Lee H, Park Y. The effects of a community-based walking program on walking ability and fall-related self-efficacy of chronic stroke patients. *Journal of exercise rehabilitation*. 2019;15(1):20-25. doi: 10.12965/jer.1836502.251.

66. Langhammer B, Stanghelle JK. Exercise on a treadmill or walking outdoors? A randomized controlled trial comparing effectiveness of two walking exercise programmes late after stroke. *Clin Rehabil*. 2010;24(1):46-54. doi: 10.1177/0269215509343328.

67. Ivey FM, Stookey AD, Hafer-Macko CE, Ryan AS, Macko RF. Higher Treadmill Training Intensity to Address Functional Aerobic Impairment after Stroke. *J Stroke Cerebrovasc Dis*. 2015;24(11):2539-2546. doi: 10.1016/j.jstrokecerebrovasdis.2015.07.002 [doi].

68. Macko RF, Ivey FM, Forrester LW, et al. Treadmill exercise rehabilitation improves ambulatory function and cardiovascular fitness in patients with chronic stroke: a randomized, controlled trial. *Stroke*. 2005;36(10):2206-2211. doi: 10.1161/01.STR.0000181076.91805.89.