Figure S1. Overall Study Diagram.

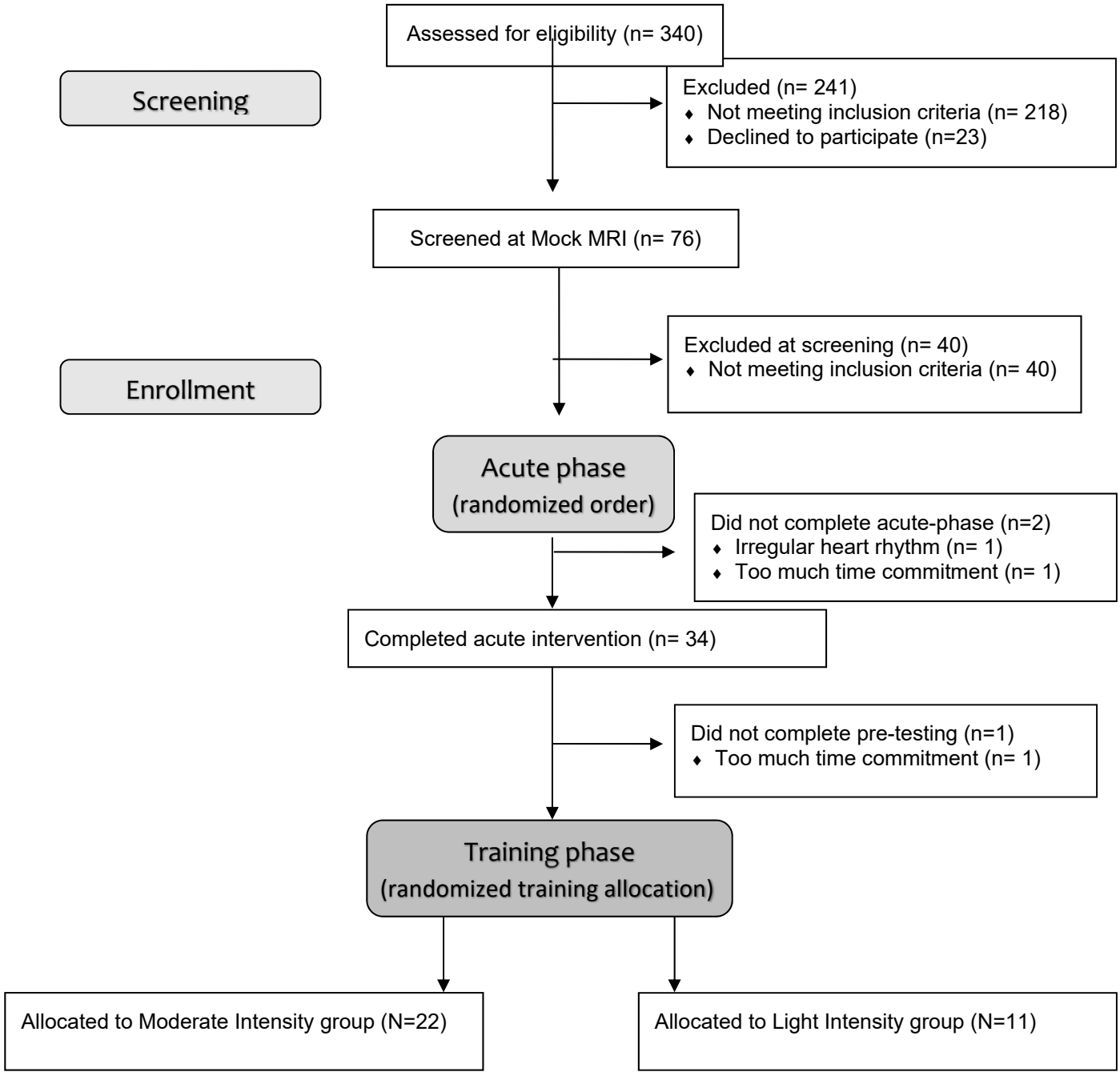

Table S1 Regions of Interest

Posterior-medial (PM) system ROIs

PM ROIs	x	y	z	Anatomical description
PTHAL4	22	-30	6	Posterior thalamus
MOCC1	-2	-78	-2	Medial posterior occipital cortex
OCCP1	-16	-96	22	Occipital pole
OCC2	16	-98	20	Occipital cortex
PREC3	18	-68	24	Precuneus
MOCC3	14	-72	8	Medial posterior occipital cortex
OCCP2	14	-88	4	Occipital pole
PHIPP1	-20	-30	-8	Posterior hippocampus
PHIPP2	20	-30	-6	Posterior hippocampus
PHC1	-14	-50	-6	Parahippocampal cortex
PHC2	18	-46	-4	Parahippocampal cortex
PREC1	-14	-60	18	Precuneus
MOCC2	6	-58	14	Medial posterior occipital cortex
OCC1	-38	-82	28	Occipital cortex
ANG2	52	-48	28	Angular gyrus
PREC5	-2	-60	34	Precuneus
RSC1	-8	-50	14	Retrospenial cortex
PREC2	-12	-50	40	Precuneus
RSC2	8	-46	16	Retrospenial cortex
PREC4	18	-52	36	Precuneus

Anterior-temporal (AT) system ROIs

AT ROIs	x	y	z	Anatomical description
DLPFC1	-24	60	24	Dorsolateral prefrontal cortex
MPFC	-2	60	34	Medial prefrontal cortex
DLPFC2	18	58	24	Dorsolateral prefrontal cortex
TPC1	-44	4	-42	Temporoparietal cortex
PMTG2	54	-2	-32	Posterior middle temporal gyrus
AHIPP2	22	-4	-26	Anterior hippocampus
OFC2	-6	16	-22	Orbitofrontal cortex
AHIPP1	-24	-12	-30	Anterior hippocampus
OFC3	24	12	-24	Orbitofrontal cortex
TPC2	38	20	-40	Temporoparietal cortex

OFC1	-16	24	-20	Orbitofrontal cortex
PMTG1	-64	-36	-10	Posterior middle temporal gyrus
AITG2	64	-14	-28	Anterior inferior temporal gyrus
FPC2	38	60	-10	Frontopolar cortex
FUS2	40	-18	-28	Fusiform cortex
FPC1	-44	58	-18	Frontopolar cortex
PRC	30	-12	-36	Perirhinal cortex
FUS1	-42	-14	-30	Fusiform cortex
OFC4	8	22	-20	Orbitofrontal cortex
PMTG3	70	-38	-10	Posterior middle temporal gyrus

Ventrolateral Prefrontal Cortex (VLPFC) ROIs

VLPFC ROIs	x	y	z	Anatomical description
ANG1	-40	-48	50	Angular gyrus
AITG2	-56	-32	-24	Anterior inferior temporal gyrus
POST2	42	-46	42	Postcentral gyrus
OCCP3	-8	-58	-30	Occipital pole
IFG2	54	38	2	Inferior frontal gyrus
PSTG2	66	-44	8	Posterior superior temporal gyrus
TPJ	60	-40	34	Temporoparietal junction
VLPFC1	54	18	12	Ventrolateral prefrontal cortex
IFG3	56	26	-8	Inferior frontal gyrus
IFG1	-58	20	-2	Inferior frontal gyrus
PSTG1	-66	-36	24	Posterior superior temporal gyrus
PRE3	60	4	4	Precuneus
POST3	62	-10	16	Postcentral gyrus
POST1	-44	-36	52	Postcentral gyrus
PRE1	-34	-20	54	Lateral precentral gyrus
PRE2	-52	4	12	Lateral precentral gyrus

Self-report measures of affect during acute exercise sessions

Participants self-reported changes in affect before, during, and immediately after completing each acute exercise bout. These were established self-report measures that have good internal consistency and have been widely used in acute exercise studies: Felt Arousal Scale (FAS) (1), Feeling Scale (FS) (2), and Positive and Negative Affect Schedule (PANAS) (3). The FS and FAS measures were collected at 2-minute intervals during the acute exercise, whereas the PANAS was assessed during the waiting room, exercise mid-point, exercise cool-down, and after post-exercise scanning.

The FAS measures arousal by asking participants to rate how "worked up" they feel on a scale of 1 to 6, where 1 = low arousal (e.g., relaxation, boredom, calmness) and 6 = high arousal (e.g., excitement, anxiety, anger). As expected, participants reported being slightly more aroused during the moderate intensity condition compared to the light ($F(1,32)=14.37$, $p=.001$), with an additional condition x time interaction ($F(11,352)=2.15$, $p=.02$) driven by increased arousal at the start of exercise and a dip at the end of cooldown (see Figure S2).

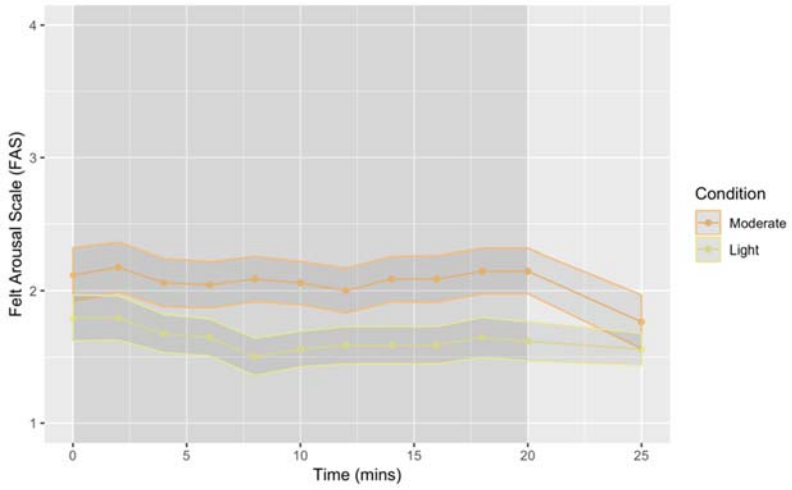

Figure S2. Change in self-reported felt arousal during acute exercise sessions.

The Feeling Scale (FS) measures affect by asking participants to rate how they feel on a scale of -5 = Very bad to +5 = Very good, with +1 shown as "Fairly good" and +3 shown as "Good" on the positive range of pleasurable feelings. Although we expected moderate intensity exercise to evoke slightly more "good" feelings, participants reported similar feelings of pleasure for each exercise condition ($F(1,33)=1.58$, $p=.22$) (see Figure S3). However, a condition x time interaction ($F(11,363)=2.78$, $p=.002$) also suggested that moderate intensity exercise tended to result in feeling better over time (Figure S3).

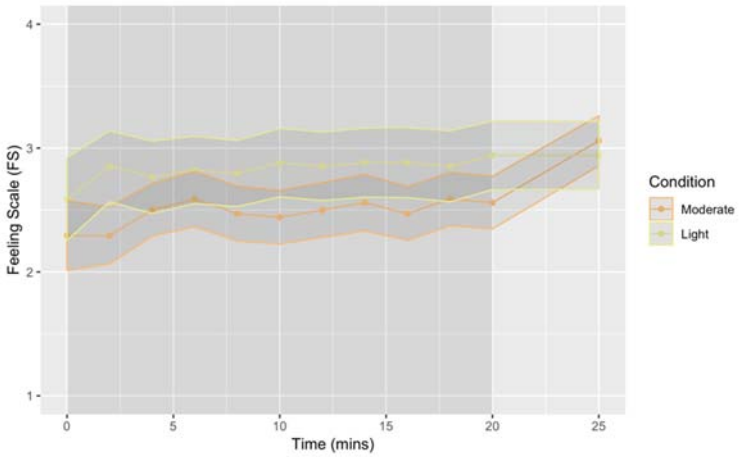

Figure S3. Change in the self-report feeling scale during acute exercise sessions.

The PANAS measures self-reported affect by having participants indicate their level of agreement to 20 words that vary on positive or negative affect, with a scale from "very slightly or not at all" (1) to "extremely" (5) for each word. Thus, positive affect scores can range from 10 to 50, with higher scores reflecting higher levels of positive affect. Negative affect scores can also range from 10 to 50, with lower scores

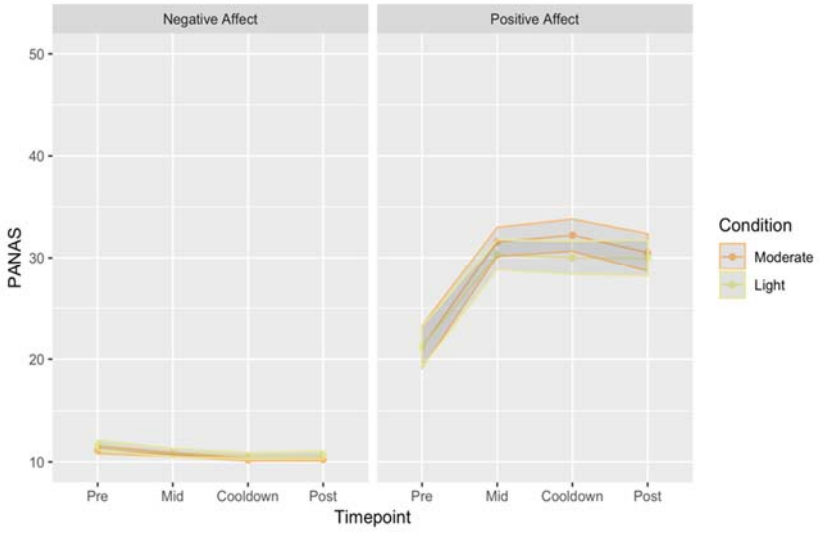

Figure S4. Change in self-reported affect on PANAS during acute exercise sessions.

reflecting lower levels of negative affect. On average, as shown in Figure S4, participants had a greater increase in positive affect in response to exercise (time x condition interaction, $F(3,87) = 52.04$, $p<.001$), but the change in positive affect was similar for both intensities (time x condition x affect interaction, $F(3,87)$ .87, $p=.46$).

Magnetic resonance imaging scanner differences did not affect longitudinal analyses

As described in the main report, due to an unexpected change in scanner, the MR images were collected on either a Siemens (SE) or General Electric (GE) 3T scanner. However, critically, all participants had both acute days and their post-intervention scans on the same machine. Therefore, the scanner change could only affect results if scanner interacted with signal quality over time. In total, 22% (2/9) of the light intensity group, and 38% (6/16) of the moderate intensity group were scanned on the SE. We examined scanner effects with linear regression models that included fixed effects for scanner (SE, GE), time (pre, post), and condition (acute: light, moderate) or group (training: light, moderate). To examine scanner effects over space and time in the fMRI signal, we focused on quality control metrics computed by MRIQC (4) that characterized spatial signal to noise (SNR, mean average BOLD signal values within the head relative to the standard deviation of those values) and temporal SNR (tSNR, mean global BOLD signal over time relative to the standard deviation of those values).

For the acute phase, we examined a linear mixed model with fixed effects (each effect coded) of scanner, condition (light, moderate), and time (pre, post), and their three-way interaction, with subject as a random effect. Results showed that across all sessions the average SNR was 4.37 (SE=.07), with a main effect of scanner (b = -.38(.13), $t(32)=-2.95$, $p<.05$) that favored the GE,

but no interactions between scanner and condition or time (p' s> .05). For tSNR, across all sessions the average tSNR was 51.82 (SE=2.38), with a main effect of scanner (b = -16.93(4.76), t(32)=-3.56, p<.05), but no interactions between scanner and condition or time (p's > .05). Therefore, as expected the newer scanner had better signal quality, but this did not interact with condition or time.

For the training phase, we examined a linear mixed model with fixed effects (each effect coded) of scanner, training group (light, moderate), and time (pre, post), and their three-way interaction, with subject as a random effect. Results showed that across all sessions the average SNR was 4.39 (SE=.08), with this time no main effect of scanner (b = -.27(.17), t(29)=-1.61, p=.12), and no interactions between scanner and condition or time (p's>.05). For tSNR, across all sessions the average tSNR was 54.99 (SE=2.40), with a main effect of scanner (b = -12.28(4.79), t(27)=-2.57, p<.05), but no interactions between scanner and condition or time (p's>.05). Therefore, results are similar to the acute phase, where although the signal quality was better for the GE scanner, this did not interact with condition or time.

Stability of motion within subjects

Frame-wise displacement (FD) was highly stable through imaging sessions that were within-day (e.g., $light_{pre}$ vs $light_{post}$ r = .89, $moderate_{pre}$ vs $moderate_{post}$ r = .93) and across days (e.g., $moderate_{pre}$ vs $light_{pre}$ r = .88), with high correlations observed to fall on the identity line. Linear mixed-effects models also showed no evidence of change in motion across the acute (b=-.01 (.02), t(33)=-.60, p=.56) or pre-to-post training (b=-.02 (.01), t(31)=-1.5, p=.14) sessions.

Lifestyle physical activity did not change from the intervention

All participants wore a GT9X ActiGraph Link accelerometer (ActiGraph; Pensacola, Florida) on the non-dominant wrist for seven consecutive days during both wake and sleep. Participants were asked to wear the device at all times except during water activities (e.g., showering, washing dishes, etc.). They wore the monitor once before randomization and again during the last week of training. A daily activity log was maintained to cross-check times in and out of bed, non-wear times, and perceived exertion. Data were collected at 60 Hz, processed in 60-second epochs with the ActiLife v6.13.3 software. Only days when the device was worn 90% of wake time were considered valid, and a minimum of 3 valid days (2 week-days + 1 weekend day) was required for inclusion in analyses. Sleep epochs were determined automatically with the Cole-Kripke (5) and checked and edited manually based on the activity log. The R package GGIR (6, 7) was used to estimate daily minutes of moderate-to-vigorous activity (MVPA). Time spent in sedentary and non-sedentary time was estimated using an R script developed by the Freedson research group at the University of Massachusetts (8) modified for 60 Hz data. Time spent in sleep, sedentary, MVPA, and non-wear were then used to assign remaining time to light intensity activity. For all activity variables average minutes per day were determined as total minutes divided by the number of valid days. Descriptive summaries for PA and sleep are shown in Table S2. Longitudinal linear models testing for an interaction between group and session, with covariates of age and sex, showed that none of the PA variables showed differential change by group (all p's > .05). Because the structured intervention was on a recumbent bicycle, lab training sessions were not expected to affect MVPA minutes based on wrist-worn accelerometry. Overall, accelerometer results suggest that participants did not change their lifestyle PA outside the lab, and that this pattern was similar for both exercise groups.

Table S2 Physical Activity before and after the intervention

	Sedentary (min/day)	Light (min/day)	Moderate-to-Vigorous (min/day)*	Sleep (min/day)
Light+				
pre	617.4 (83.7)	303.7 (89.5)	48.0 (29.6)	405.7 (54.3)
post	654.7 (99.5)	281.0 (62.6)	46.9 (35.4)	408.2 (65.0)
Moderate				
pre	646.4 (104.1)	271.1 (79.0)	36.3 (31.5)	436.9 (51.4)
post	648.5 (101.6)	274.8 (76.7)	24.9 (54.4)	434.5 (60.1)

*Moderate-to-vigorous is median (IQR) because it is positively skewed; groups were not statistically significantly different in PA or sleep before the intervention.

Detailed description of statistical model for testing acute and training effects, and their residualized change scores, for functional connectivity and working memory

Linear mixed models (LMMs) were used to test whether acute changes in memory system *fc* and working memory performance predict changes in the same outcomes after training. Models adjusted for age and sex, and a standard *t*-test with estimated Kenward-Roger *df* (9) was used to evaluate the null hypotheses. The *p*-values for brain network analyses were adjusted for multiplicity using an optimized False Discovery Rate (FDR) approach (10), yielding *q*-values.

Let y_{ij} represent *fc* between each ROI-ROI pair for the *i*th participant ($i = 1, \ldots, N$) at the *j*th visit ($j = 1, 2$). Then the LMM testing acute change was $y_{ij} = \beta_0 + \beta_1 time_{ij} + \beta_2 condition_i + \beta_3(time_{ij} \cdot condition_i) + \beta_4 sex_i + \beta_5 age_i + b_{0i} + b_{1i} time_{ij} + b_{3i}(time_{ij} \cdot condition_i) + e_{ij}$, where β_k is a fixed effect, b_{ki} is a random effect, and e_{ij} is random error. In the above equation, $condition_i$ takes the value of 1 for the active condition and 0 for passive; $time_{ij}$ takes

the value of 1 for (post) and the value of 0 for pre-test sessions; sex_i takes the value of 1 for female and 0 for male; and age_i takes the value of 1 for the block 70-80 and 0 otherwise.

An acute score was then derived from the parameter estimates of the acute phase LMM, which included estimated randomized effects for each outcome for each participant. Fixing sex and age, the binary coding of the variables in Equation 1 indicate the acute score can be computed as: Acute score = [($\text{moderate}_{\text{post}} - \text{moderate}_{\text{pre}}$) – ($\text{light}_{\text{post}} - \text{light}_{\text{pre}}$)] = $\hat{\beta}_3 + \hat{b}_{3i}$. Higher values for the acute score indicated that the participant responded more to moderate compared to light intensity exercise.

In order to adjust for the fact that the day 1 pre-test session was in both acute and training models, a residual change score was defined for both the acute and training model. The acute residual change model was defined as: ($\text{moderate}_{\text{post}} - \text{moderate}_{\text{pre}}$) – ($\text{light}_{\text{post}} - \text{light}_{\text{pre}}$) = $\beta_0 + \beta_1 \text{Pre}_{\text{M/L i}} + r_i$, where $\text{Pre}_{\text{/L}}$ is the day 1 pre-test scores. r is the residualized change score for the acute model, defined as the difference of difference minus the intercept and slope of the pre-test scores. Similarly, the training residual change model was defined as the post interventions scan minus the pre-test scores, and was modeled as a function of the pre-test score: post intervention scan $- \text{Pre}_{\text{M/L}} = \beta_0 + \beta_1 \text{Pre}_{\text{M/Li}} + e_i$, where e_i is the residualized change score for the training model, defined as the difference minus the intercept and slope of the pre-test scores.

Supplemental References

1. Svebak S, Murgatroyd S. Metamotivational dominance: a multimethod validation of reversal theory constructs. Journal of Personality and Social Psychology. 1985;48(1).

2. Hardy CJ, Rejeski WJ. Not what, but how one feels: the measurement of affect during exercise. Journal of Sport and Exercise Psychology. 1989;11(3):304-17.

3. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology. 1988;54(6):1063.

4. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One. 2017;12(9):e0184661. Epub 2017/09/26. doi: 10.1371/journal.pone.0184661. PubMed PMID: 28945803; PubMed Central PMCID: PMCPMC5612458.

5. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep. 1992;15(5):461-9. Epub 1992/10/01. PubMed PMID: 1455130.

6. van Hees VT, Golubic R, Ekelund U, Brage S. Impact of study design on development and evaluation of an activity-type classifier. J Appl Physiol (1985). 2013;114(8):1042-51. Epub 2013/02/23. doi: 10.1152/japplphysiol.00984.2012. PubMed PMID: 23429872; PubMed Central PMCID: PMCPMC3633433.

7. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. PLoS One. 2015;10(11):e0142533. Epub 2015/11/17. doi: 10.1371/journal.pone.0142533. PubMed PMID: 26569414; PubMed Central PMCID: PMCPMC4646630.

8. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. J Appl Physiol (1985). 2015;119(4):396-403. Epub 2015/06/27. doi: 10.1152/japplphysiol.00026.2015. PubMed PMID: 26112238; PubMed Central PMCID: PMCPMC4538283.

9. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997:983-97.

10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological). 1995:289-300.