SLEEP, 2024, XX, 1–15
https://doi.org/10.1093/sleep/zsae250
Advance access publication 24 October 2024
Original Article

Original Article

A replicate crossover trial on the interindividual variability of sleep indices in response to acute exercise undertaken by healthy men

Yuting Yang[1,2], Alice E. Thackray[1,3], Tonghui Shen[1,3], Tareq F. Alotaibi[4,5], Turki M. Alanazi[6,7], Tom Clifford[1], Iuliana Hartescu[1], James A. King[1,3], Matthew J. Roberts[1,3], Scott A. Willis[1,3], Lorenzo Lolli[8], Greg Atkinson[9] and David J. Stensel[1,3,10,11,*]

[1]National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom,
[2]School of Sports Medicine and Rehabilitation, Beijing Sport University, Beijing, China,
[3]National Institute for Health and Care Research (NIHR) Leicester Biomedical Research Centre, University Hospitals of Leicester NHS Trust and University of Leicester, Leicester, United Kingdom,
[4]King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia,
[5]King Abdullah International Medical Research Center, Riyadh, Saudi Arabia,
[6]Department of Respiratory Therapy, College of Applied Medical Sciences, King Saud Bin Abdulaziz University for Health Sciences, Al Ahsa, Saudi Arabia,
[7]King Abdullah International Medical Research Center, Al Ahsa, Saudi Arabia,
[8]Department of Sport and Exercise Sciences, Institute of Sport, Manchester Metropolitan University, Manchester, United Kingdom,
[9]School of Sport and Exercise Science, Liverpool John Moores University, Liverpool, United Kingdom,
[10]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan and
[11]Department of Sports Science and Physical Education, The Chinese University of Hong Kong, Hong Kong

*Corresponding author. David Stensel, National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Leicestershire LE11 3TU, United Kingdom. Email: d.j.stensel@lboro.ac.uk.

Abstract

Study Objectives: Using the necessary replicate-crossover design, we investigated whether there is interindividual variability in home-assessed sleep in response to acute exercise.

Methods: Eighteen healthy men (mean [*SD*]: 26[6] years) completed two identical control (8 hour laboratory rest, 08:45–16:45) and two identical exercise (7 hour laboratory rest; 1 hour laboratory treadmill run [62(7)% peak oxygen uptake], 15:15–16:15) trials in randomized sequences. Wrist-worn actigraphy (MotionWatch 8) measured home-based sleep (total sleep time, actual wake time, sleep latency, and sleep efficiency) two nights before (nights 1 and 2) and three nights after (nights 3–5) the exercise/control day. Pearson's correlation coefficients quantified the consistency of individual differences between the replicates of control-adjusted exercise responses to explore: (1) immediate (night 3 minus night 2); (2) delayed (night 5 minus night 2); and (3) overall (average post-intervention minus average pre-intervention) exercise-related effects. Within-participant linear mixed models and a random-effects between-participant meta-analysis estimated participant-by-trial response heterogeneity.

Results: For all comparisons and sleep outcomes, the between-replicate correlations were nonsignificant, ranging from trivial to moderate (r range = –0.44 to 0.41, $p \geq .065$). Participant-by-trial interactions were trivial. Individual differences *SD*s were small, prone to uncertainty around the estimates indicated by wide 95% confidence intervals, and did not provide support for true individual response heterogeneity. Meta-analyses of the between-participant, replicate-averaged condition effect revealed that, again, heterogeneity (τ) was negligible for most sleep outcomes.

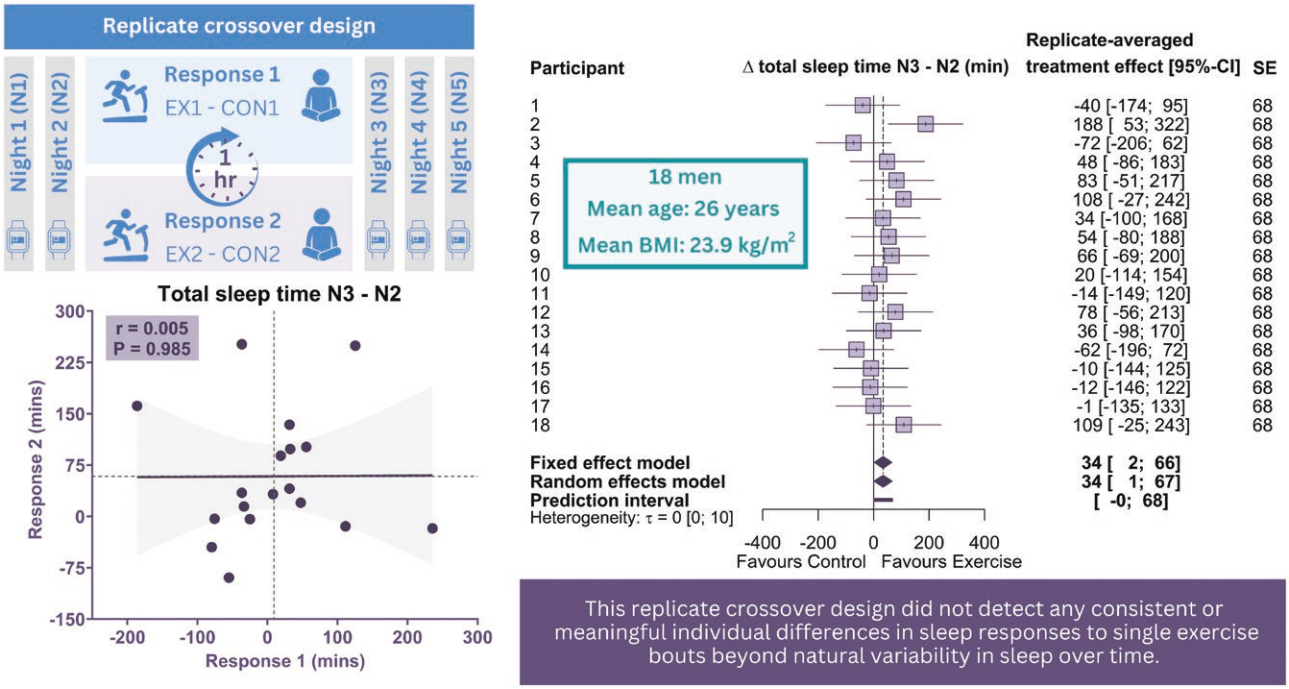
Conclusions: Control-adjusted sleep in response to acute exercise was inconsistent when measured on repeated occasions. Interindividual differences in sleep in response to exercise were small compared with the natural (trial-to-trial) within-subject variability in sleep outcomes.

Clinical trials information: https://clinicaltrials.gov/study/NCT05022498. Registration number: NCT05022498.

Key words: actigraphy; sleep; exercise; individual variability; replicate crossover

Submitted for publication: August 22, 2024; **Revised:** October 14, 2024

Graphical Abstract



Statement of Significance

Single exercise bouts have been shown to elicit modest but positive effects on sleep, including improvements in sleep duration and sleep quality, but the magnitude of true individual response heterogeneity has not been quantified. We adopted a robust and novel study design, the replicate crossover, and associated statistical approaches to quantify the consistency and magnitude of interindividual variability in home-based, actigraphy-derived sleep in response to an afternoon bout of treadmill running. Our findings revealed substantial trial-to-trial within-subject variability which inhibited the detection of any consistent and meaningful exercise-related response heterogeneity in sleep outcomes. Inferences on interindividual variability in sleep responses to an (exercise) intervention may be misleading if patient-by-treatment response heterogeneity is not appropriately separated from other sources of variability.

Sleep is crucial for preserving physiological and cognitive function across the lifespan [1]. Accumulating an average of 7–9 hours of sleep per night is recommended for most adults [2, 3] but, despite its biological necessity, many adults fail to achieve sufficient duration and/or quality of sleep for health [4, 5]. This presents a major public health concern given that curtailed (<7 hours/night), excess (>9 hours/night), and/or disrupted sleep has been associated with a myriad of adverse health consequences, including obesity, type 2 diabetes, cardiovascular disease, and all-cause mortality [6–8]. Furthermore, accumulating evidence suggests that achieving consistent day-to-day sleep patterns is important for enhancing health [9], with sleep regularity identified as a stronger predictor of all-cause mortality than sleep duration [10].

Physical activity is widely promoted as a nonpharmacological strategy for enhancing sleep [11]. Evidence from observational work has demonstrated that higher levels of physical activity may mitigate the elevated cardiometabolic health and mortality risk associated with poor sleep patterns [12, 13]. Comprehensive reviews suggest that acute (single bouts) and short-term exercise (<1 week in duration) performed at any time of day elicits modest but positive effects on sleep, including increasing total sleep time and sleep efficiency, reducing sleep onset latency and improving aspects of sleep architecture [14, 15]. Furthermore, contrary to sleep hygiene recommendations advocating that exercise proximal to bedtime is detrimental to sleep quality [16], the consensus of evidence suggests that single bouts of evening physical activity or structured exercise completed ~15 minutes to 4 hours before bedtime do not disrupt sleep during the subsequent night [17–19].

The notion that individual differences exist in response to the same intervention has become a popular avenue of inquiry within the context of *precision medicine* [20]. However, reliable inferences can only be obtained with appropriate research designs and statistical models [21, 22]. One such approach is the replicate crossover design, which involves repeated administration of control and intervention conditions to quantify the participant-by-treatment interaction defined as the extent that the treatment response differs between participants [23–25]. This design allows the estimation of various components of variation to examine whether genuine treatment response heterogeneity can be distinguished from random within-subject variability allowing researchers to arrive at appropriate conclusions on the presence and consistency of interindividual differences [21–24]. Several studies have adopted this framework with associated statistical approaches [21–27] to identify true interindividual variability in appetite responses to acute exercise and standardized meal intake [28, 29],

and blood pressure responses to antihypertensive medications [30]. Conversely, adopting a similar design and analysis approach, genuine interindividual heterogeneity could not be identified in cardiovascular disease risk marker responses to acute exercise [31], highlighting the importance of accounting for trial-to-trial within-subject variability and measurement error [21–24]. Existing work suggests substantial night-to-night variability is evident in subjective perceptions of sleep and objective measures of sleep duration and quality [32–35]. However, it remains unknown whether there is true interindividual variability in sleep in response to acute exercise beyond random within-subject variability over time.

This study examined the consistency of home-assessed sleep using actigraphy (total sleep time, actual wake time, sleep latency, and sleep efficiency) in response to acute bouts of moderate-intensity exercise and quantified the magnitude of interindividual differences in responses using a replicate crossover research design. It was hypothesized that exercise-induced changes in sleep outcomes would be consistent on repeated occasions, and true interindividual variability in exercise-related sleep would be observed in young healthy men.

Materials and Methods

Ethical approval and participants

This manuscript presents secondary outcomes from a replicate crossover trial investigating interindividual variability in postprandial cardiovascular disease risk marker responses to acute exercise (ClinicalTrials.gov identifier: NCT05022498). No changes to the study methods or outcomes were made after the study commenced. A detailed description of the study protocol and primary study results have been published previously [31]. The study received approval from the Loughborough University Ethics Advisory Committee (R19-P103) before any study-related measures commenced. Twenty healthy young men were recruited from the local community and provided written informed consent to participate in the study at Loughborough University. All participants were nonsmokers, reported being weight stable (defined as ≤3 kg change in body mass in the previous 3 months), had no diagnosed cardiometabolic diseases or sleep disorders, and were not taking any medications. The recruitment of men for the study was based on recognition that postprandial triacylglycerol concentrations, the primary outcome of the study [31], are typically greater in men than premenopausal women [36] and fluctuate across the menstrual cycle [37]. Study methods and results are reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement extended to randomized crossover trials [38].

Preliminary measures

During a preliminary visit, participants completed questionnaires to assess health status, anthropometric measurements (stature, body mass, and body fat percentage) and completed treadmill familiarization (RUN RACE, Technogym, Gambettola, Italy) and exercise testing. The latter comprised a treadmill-based 16 minute submaximal incremental test and a ramped (+1% gradient each minute) peak oxygen uptake test as described previously [31], with the data used to identify the treadmill speed estimated to predict 60% of peak oxygen uptake for the main experimental trials. Throughout both tests, expired air samples were sampled continuously for oxygen consumption using an online breath-by-breath gas analyzer (MetaMax 3B; Cortex, Leipzig, Germany), and peak oxygen uptake was calculated as the highest 30 second rolling average. Heart rate (Polar T31; Polar Electro, Kempele, Finland) and ratings of perceived exertion (Borg CR-10 [39]) were also measured throughout.

Study design

Using a replicate crossover design [24], participants completed two identical exercise and two identical control trials in a randomized order separated by at least 5 days. The randomization sequence for the four experimental trials was obtained from an online software tool (http://www.sealedenvelope.com/) by a researcher who was not involved directly in data collection (G.A.). Other researchers (T.S., T.F.A., T.M.A.) enrolled participants and assigned participants to the sequence of interventions. Main trials involved four 2 day (days 1 and 2) laboratory visits, but the analysis presented in this paper involves sleep assessments collected during the two nights before day 1 (nights 1 and 2), the night of day 1 (night 3), and two nights after day 1 (nights 4 and 5). Participants were informed of the trial allocation upon arrival at the laboratory on day 1 of each main trial. Alcohol, caffeine, and strenuous physical activity (outside the laboratory) were not permitted for 24 hours before day 1 or until the laboratory measures were completed at 16:45 on day 2 of each main trial. A weighed dietary record was completed in the 48 hours before the first main trial day which was replicated in the same period before subsequent trials. Adherence to the standardization procedures was confirmed verbally upon attendance at the laboratory. A schematic of the trial protocol is shown in Figure 1.

Main trials

On day 1, participants arrived at the laboratory at 08:00 after fasting overnight for 10 hours. Participants rested in the laboratory until 16:45 throughout all trials apart from performing 1 hour of treadmill running (1% gradient, 60% peak oxygen uptake) between 15:15 and 16:15 in the two exercise trials. Expired air samples were measured continuously to monitor exercise intensity and estimate exercise energy expenditure and substrate utilization [40]. Heart rate and ratings of perceived exertion were recorded at 10 minute intervals, and the treadmill speed was adjusted periodically to ensure that the target intensity was achieved.

Standardized breakfast and lunch meals were consumed within 10 minutes at ~08:35 and ~12:45, respectively, and plain water was provided ad libitum during the trials. Breakfast consisted of plain croissants, milk chocolate spread, double cream, and chocolate milkshake, providing 60 kJ energy per kilogram of body mass (57% fat, 35% carbohydrate, 8% protein). Lunch consisted of white bread, Cheddar cheese, butter, double cream, and chocolate milkshake, which provided 60 kJ energy per kilogram of body mass (60% fat, 28% carbohydrate, 12% protein). Participants were provided with a standardized evening meal (Margherita pizza; 2511 kJ, 32% fat, 52% carbohydrate, 16% protein) and were asked to consume the meal before 22:00. After the meal, only plain water was allowed until the participants arrived at the laboratory the next morning (i.e. day 2) at 08:00 where they rested throughout the day in all experimental trials until 16:45 and consumed the same standardized breakfast and lunch meals at 08:35 and 12:45, respectively.

Sleep assessment

Habitual sleep was assessed using a wrist-worn actigraphy device (MotionWatch 8; CamNTech, Cambridge, UK) in combination with

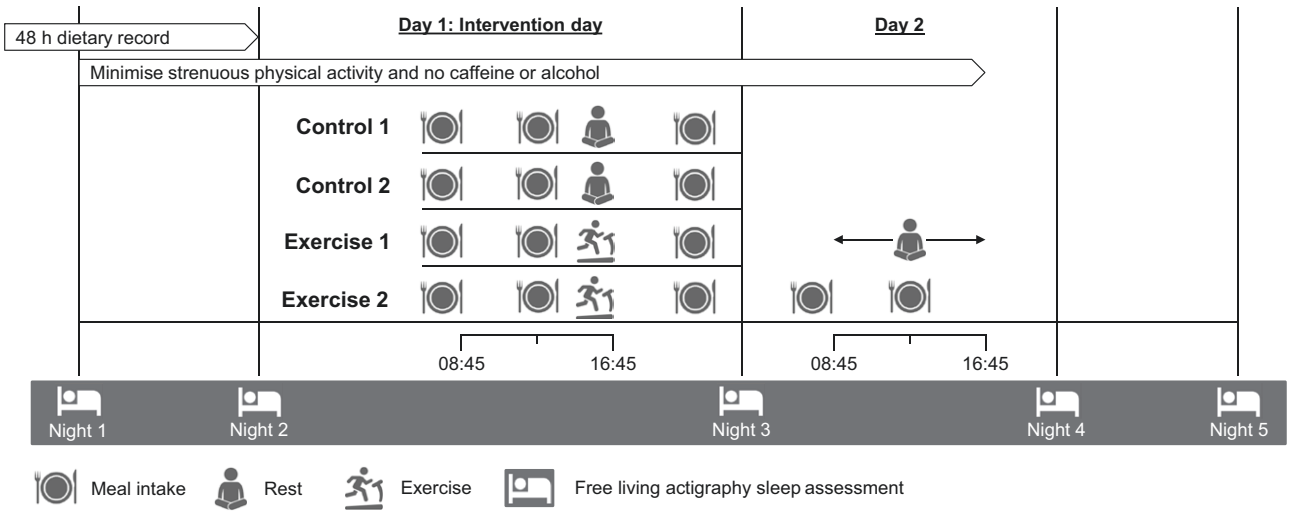


Figure 1. Schematic of the study protocol. Experimental trials were completed in a randomized order separated by at least 5 days.

a sleep diary. Participants wore the watch on their nondominant wrist for five consecutive nights in each trial to capture sleep during: (i) the two nights before day 1 of the main trials (nights 1 and 2); (ii) the same night of the exercise and rest interventions on day 1 of the main trials (night 3); and (iii) the two nights after day 1 of the main trials (nights 4 and 5) (Figure 1). Participants were instructed to follow their usual sleep routine with no restrictions on sleep time, location, or screen use before sleep. Due to collecting five consecutive nights of sleep data on four repeat occasions, the start day of data collection in each trial could not be controlled either within or between participants. The MotionWatch 8 was initialized to record in tri-axial mode 1 and raw data were analyzed in 30 second epochs (MotionWare version 1.2.5; CamNTech, Cambridge, UK) to derive four sleep variables: total sleep time (actual time spent in sleep), actual wake time (time awake after sleep onset), sleep latency (time between lights out and sleep onset) and sleep efficiency (total sleep time expressed as a percentage of time in bed). The device uses an algorithm validated against the criterion polysomnography method for sleep detection [41, 42] (further details of the sleep scoring process are provided in the Supplementary Methods). Participants were asked to press the event marker on the actigraphy device at sleep and wake times, which was used in conjunction with a sleep diary documenting daily sleep time, awakenings, and sleep quality to help identify sleep time.

Statistical analyses

The data reported in this manuscript are secondary outcomes from a previous study [31] where a pragmatic approach to sample size justification was adopted due to the onerous nature of the study design (for further details, see 31). The statistical analysis framework consisted of a four-step approach in line with existing research [23, 24, 28, 29, 31] and more recent advances [43] for the appropriate examination of continuous data from a replicate crossover experiment. All statistical analysis was performed upon study completion once the target sample had completed the study requisites. Response pairs were formulated by calculating the control-adjusted treatment effect for the first exercise and control pair in each participant's sequence (response 1; exercise 1 minus control 1) along with the second exercise and control pair (response 2; exercise 2 minus control 2). The first and second replicates were calculated as the difference between the exercise and control trial pre-to-post change scores in three comparisons:

(1) The immediate impact of exercise was explored by calculating the pre-to-post change score of night 3 minus night 2;
(2) The delayed effect of exercise was explored by calculating the pre-to-post change score of night 5 minus night 2; and
(3) The overall influence of exercise was explored by calculating the pre-to-post change score of the average of nights 3–5 (post-intervention) minus the average of nights 1 and 2 (pre-intervention).

First, we calculated Pearson's product-moment correlation coefficients (*r*) between the two replicates of the exercise and control pre-to-post change for each sleep outcome [24]. This provides an indication of the consistency of the exercise effect between the replicate trials, with thresholds of 0.1, 0.3, and 0.5 indicating small, moderate, and large coefficients, respectively [44].

Second, a naïve estimate (estimate 1) of the *SD* for individual responses (SD_{IR}) was determined by calculating the difference in *SD*s of the pre-to-post change between the exercise and control trials as follows:

$$SD_{IR} = \sqrt{SD_E^2 - SD_C^2}$$

where the SD_{IR} is the *SD* of the true (control adjusted) individual variability in the exercise effect, and SD_E and SD_C are the *SD*s of the pre-to-post change in the exercise and control trials, respectively [21, 22, 26]. These calculations utilized the appropriate equation for pooling *SD*s across the two replicates of the exercise and control trials [45]. A positive SD_{IR} indicates greater heterogeneity in the exercise response compared with any random within-subject variability.

Given the above naïve estimation of the SD_{IR} (estimate 1) was originally designed to compare variances between groups in parallel arm trials rather than between conditions in crossover trials, we also employed within-participant linear mixed models, conducted in SAS OnDemand for Academics [46]. These models derived the participant-by-trial interaction for each sleep outcome and comparison by modeling trial, period (trial sequence), and the period-by-trial interaction as fixed effects, whereas participant and the participant-by-trial interaction were modeled as random effects (see SAS base code presented

in Supplementary Methods). Estimate 2 of the true individual differences SD was derived from the participant-by-trial interaction term. Standard residual diagnostics were performed to assess the adequacy and stability of the modeled covariance parameter estimates [47, 48], and a sensitivity analysis was conducted excluding outliers that were >3 times higher or lower than the sample SD.

Fourth, we calculated a sample estimate of within-subjects variance and converted it to a standard error (*SE*) using appropriate degrees of freedom given the completed cycles to derive per participant replicate-averaged treatment effects [43]. A random-effects meta-analysis with Hartung–Knapp adjustment [49] summarized individual-participant, replicate-averaged treatment effects, and respective sampling errors [43] conducted using the *metagen()* function [50]. Because our study design was a fully "balanced" replicate crossover design rather than an *n*-of-1 trial involving a varying number of condition repeats between participants, a consistent *SE*, and 95% confidence interval (CI), for all participants resulted. Senn [43] provides more details about this *SE* component of the meta-analysis approach. The restricted maximum-likelihood estimation method determined the tau-statistic (τ) value describing the between-participant, replicate-averaged treatment effect response variability across the distribution of true acute exercise-related effects [51, 52]. The uncertainty surrounding the point τ-statistic estimate was described using 95% CIs derived using the generalized Q-statistic method [53]. Weighted raw replicate-averaged treatment effects were reported as descriptive statistics alongside the respective 95% prediction interval illustrating the range for the distribution of true mean differences expected for 95% of similar trials [54, 55]. Meta-analyses were conducted in R (version 4.2.2, R Foundation for Statistical Computing; see R base code in Supplementary Methods).

Mean differences and correlation coefficients are reported with their corresponding 95% CI. In the absence of a clinical anchor to ascertain meaningful between-trial differences in sleep, absolute standardized effect sizes (Cohen's *d*) were calculated [56] and thresholds of 0.2, 0.5, and 0.8 were considered small, moderate, and large effects, respectively [44]. A minimal clinical important difference (MCID) of ±30 minutes was adopted for total sleep time, actual wake time, and sleep latency, and ±5% was adopted for sleep efficiency. These values are in line with proposed thresholds defining the acceptable difference in sleep outcomes between actigraphy and polysomnography devices [57, 58] and are broadly consistent with the expected night-to-night variability in these outcomes [32, 33]. An alpha value of $p < .05$ was considered statistically significant.

Results

Participants and sleep data

Eligible participants were recruited to take part in the study between October 2019 and July 2021. Study completion was determined once the target sample size was reached, and there were no adverse events to report relating to the study protocol. Due to technical issues with the actigraphy device, the sleep data in this manuscript are reported for 18 participants. The characteristics of participants ($n = 18$) were as follows (mean [SD]): age 26 (6) years; body mass 76.3 (9.2) kg; body mass index 23.9 (2.5) kg/m^2; body fat percentage 19.3 (7.2)%; and peak oxygen uptake 44 (10) mL/kg/min.

Most participants provided sleep data on either all weekday nights or all weekend nights for night 3 (immediate exercise effect; $n = 14$, 78%) and night 5 (delayed exercise effect; $n = 15$, 83%) across all four trials, but for some participants (≤22%), these nights fell on a combination of weekday and weekend nights across the trials. Two participants had missing sleep data on night 5 in one of the experimental trials; therefore, the correlation coefficients for comparison 2 (delayed exercise effect) are presented for $n = 16$ with 32 replicates of the control-adjusted exercise response, and the modeling approaches are presented for $n = 18$ with 70 observations/trials. The flow of participants through the study is presented in Supplementary Figure 1.

Exercise responses

The 95% CIs for the mean difference in exercise responses on the two occasions overlapped zero (all $p \geq .168$), and the standardized effect sizes were trivial ($d = 0.01$–0.19) apart from the small effect size for fat oxidation ($d = 0.27$) (Table 1).

Consistency and individual variability in sleep in response to exercise

Comparison 1: night 3 minus night 2

Pearson's correlation coefficients between the two replicates of control-adjusted exercise responses were not statistically significant and ranged in magnitude from trivial (total sleep time: $r = 0.005$ [95% CI = −0.46 to 0.47], $p = .985$) to small-to-moderate (actual wake time: $r = 0.36$ [95% CI = −0.13 to 0.71], $p = .143$; sleep latency: $r = -0.44$ [95% CI = −0.75 to 0.03], $p = .065$; and sleep efficiency: $r = -0.14$ [95% CI = −0.57 to 0.35], $p = .587$) (Figure 2). For all sleep outcomes, the 95% CI for the period-adjusted mean difference between exercise and control trials overlapped zero with small-to-moderate standardized effect sizes (all main effects

Table 1. Treadmill exercise responses

Variable	Exercise trial 1	Exercise trial 2	95% CI*	Cohen's *d*	*P*-value
Oxygen uptake (L/min)	2.07 (0.49)	2.08 (0.48)	−0.07, 0.06	0.01	.850
Peak oxygen uptake (%)	62 (7)	62 (7)	−2, 2	0.03	.821
Heart rate (bpm)	156 (11)	156 (10)	−7, 7	0.01	.986
Rating of perceived exertion	6 (1)	6 (1)	−1, 1	0.07	.842
Respiratory exchange ratio	0.96 (0.04)	0.97 (0.06)	−0.03, 0.02	0.07	.750
Fat oxidation (g)	9.7 (6.5)	11.5 (7.4)	−4.6, 0.9	0.27	.168
Carbohydrate oxidation (g)	157.3 (34.2)	159.5 (44.7)	−17.1, 12.6	0.06	.755
Gross energy expenditure (kJ)	2895 (520)	3003 (593)	−274, 58	0.19	.188

All values are mean (*SD*) for $n = 18$ healthy men.
*95% CI of the mean absolute difference between the exercise trials derived from a paired sample *t*-test.

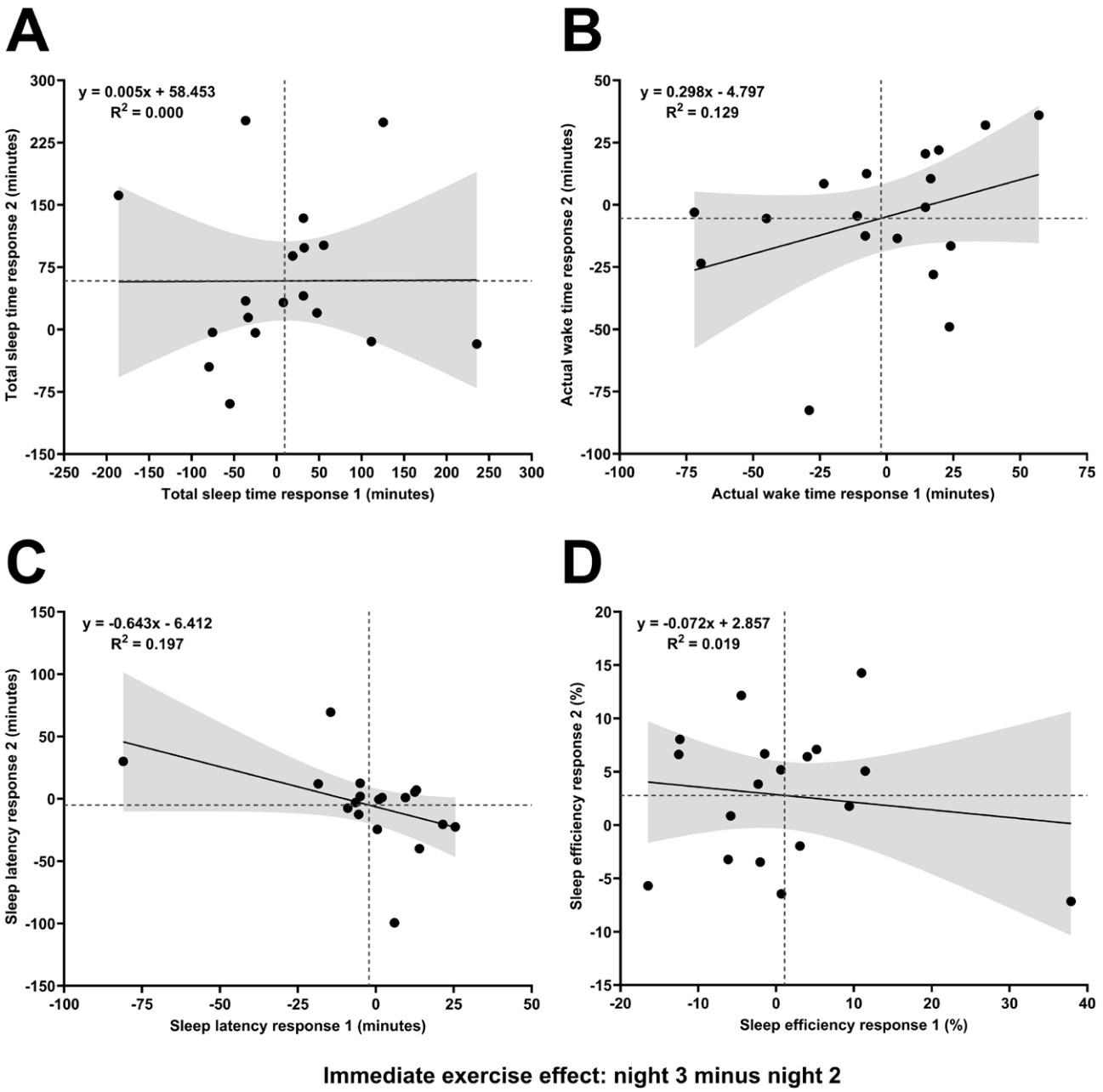


Figure 2. Relationship between exercise and control pre-to-post change scores (night 3 minus night 2) on the two occasions for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 18 men (36 replicates of the control-adjusted exercise response). "Response 1" calculated from the first pair of trials (exercise 1 minus control 1) and "Response 2" calculated from the second pair of trials (exercise 2 minus control 2). Dashed lines depict the mean responses and gray-shaded region represents the 95% CI of the regression line (black solid line).

of trial $p \geq .111$, $d = 0.34–0.55$), and participant-by-trial interactions were not statistically significant (all $p \geq .157$) (Table 2). The SD_{IR} for estimate 1 and estimate 2 was small with wide 95% CIs for all outcomes, and the SD_{IR} (estimate 2) for sleep latency and sleep efficiency was negative (Table 2). The τ statistic from the random-effects meta-analysis provides some evidence of between-participant variability for actual wake time (τ statistic [95% CI]: 19 [9, 35] minutes), but corresponding values for the other sleep outcomes were zero (Figure 3). Inspection of the individual data plots revealed that the proportion of participants exhibiting the mean control-adjusted exercise response beyond the MCID was 56% above and 17% below for total sleep time, 11% above and 17% below for actual wake time, 0% above and 6% below for sleep latency, and 33% above and 6% below for sleep efficiency (Figure 3).

Outliers were identified for five trials ($n = 5$ participants) for actual wake time, seven trials ($n = 6$ participants) for sleep latency, and three trials ($n = 3$ participants) for sleep efficiency. A sensitivity analysis removing outliers yielded similar small-to-moderate correlations for actual wake time ($r = 0.24$ [95% CI = –0.36 to 0.70]), sleep latency ($r = –0.34$ [95% CI = –0.76 to 0.29]), and sleep efficiency ($r = –0.12$ [95% CI = –0.60 to 0.42]) that were not statistically significant (all $p \geq .285$). The exclusion of these data did not markedly alter the interpretation of the participant-by-trial interactions (all $p \geq .340$), the SD_{IR} for estimate 1 or 2, or the estimated between-participant heterogeneity

Table 2. Estimated marginal means and SEs of the pre-to-post change scores for sleep outcomes in the exercise and control conditions with the true individual differences *SD*

Sleep variable	N (obs)	Exercise change, Mean (*SE*)	Control change, Mean (*SE*)	Main effect of trial*			Estimate 1†	Estimate 2‡	
				Mean difference (95% CI)	Cohen's *d*	*P*-value	Individual differences *SD*	Individual differences *SD* (95% CI)	*P*-value
Comparison 1: night 3 minus night 2									
Total sleep time (min)	18 (72)	39.5 (15.3)	5.6 (11.1)	34.0 (−8.4, 76.3)	0.51	0.111	48.0	47.1 (−40.0, 77.7)	0.255
Actual wake time (min)	18 (72)	1.7 (3.8)	8.4 (3.8)	−6.7 (−18.0, 4.5)	0.34	0.235	−12.3	4.8 (−12.9, 14.5)	0.812
Sleep latency (min)	18 (72)	−6.0 (3.0)	−1.8 (3.9)	−4.2 (−14.6, 6.2)	0.55	0.420	−12.3	−9.1 (−16.3, 10.0)	0.376
Sleep efficiency (%)	18 (72)	2.3 (0.9)	−0.5 (1.3)	2.7 (−0.7, 6.1)	0.52	0.116	−3.2	−3.5 (−5.4, 2.2)	0.157
Comparison 2: night 5 minus night 2									
Total sleep time (min)	18 (70)	22.6 (18.0)	28.2 (15.8)	−5.5 (−47.3, 36.3)	0.08	0.788	44.4	39.7 (−49.5, 74.8)	0.443
Actual wake time (min)	18 (70)	13.8 (5.3)	15.3 (6.8)	−1.5 (−15.2, 12.2)	0.08	0.823	7.4	−16.5 (−26.0, 11.6)	0.190
Sleep latency (min)	18 (70)	−3.5 (2.7)	−1.2 (4.4)	−2.3 (−10.1, 5.5)	0.35	0.545	−5.7	−13.8 (−18.1, −7.2)	0.007
Sleep efficiency (%)	18 (70)	−2.0 (1.3)	−2.4 (2.1)	0.4 (−3.0, 3.8)	0.09	0.810	4.7	−6.4 (−8.1, −3.8)	0.002
Comparison 3: average post-intervention (nights 3–5) minus average pre-intervention (nights 1 and 2)									
Total sleep time (min)	18 (72)	21.9 (9.7)	7.5 (8.9)	14.4 (−14.8, 43.7)	0.27	0.322	26.4	19.9 (−33.6, 43.8)	0.612
Actual wake time (min)	18 (72)	6.7 (3.6)	3.7 (4.0)	3.0 (−6.6, 12.6)	0.17	0.530	−5.6	−5.0 (−14.1, 12.2)	0.776
Sleep latency (min)	18 (72)	−4.4 (3.1)	−5.7 (2.9)	1.3 (−6.2, 8.8)	0.12	0.725	−7.3	5.8 (−9.5, 12.5)	0.596
Sleep efficiency (%)	18 (72)	0.2 (0.9)	0.5 (1.1)	−0.4 (−2.5, 1.8)	0.06	0.736	−1.7	−2.1 (−3.8, 2.5)	0.412

N (obs) refers to the number of participants (*N*) and observations/trials (obs) included in the analysis for each comparison and outcome.
*Estimated from a within-participant random effects linear mixed model.
†Estimate 1: Individual differences *SD* estimated using $SD_{IR} = \sqrt{SD_E^2 - SD_C^2}$ where SD_{IR} is the *SD* of the true individual response, and SD_E and SD_C are the SDs of the pre-to-post change scores for the exercise and control trials (averaged over both replicates), respectively [21].
‡Estimate 2: Period-adjusted individual differences *SD* estimated using a random effects statistical model [23]. The *SD* was calculated from the participant-by-trial interaction term modeled as a random effect (refer to the SAS code supplied in Supplementary Methods). The *P*-value shown is also for this interaction term.

from the random effects meta-analysis (τ range 0–15 minutes) (Supplementary Table 1).

Comparison 2: night 5 minus night 2.

Small but statistically nonsignificant Pearson's correlation coefficients were identified between the two replicates of control-adjusted exercise responses for total sleep time ($r = 0.16$ [95% CI = −0.37 to 0.61], $p = .563$), actual wake time ($r = -0.12$ [95% CI = −0.58 to 0.40], $p = .660$), sleep latency ($r = -0.23$ [95% CI = −0.65 to 0.30], $p = .395$), and sleep efficiency ($r = -0.10$ [95% CI = −0.57 to 0.42], $p = .719$) (Figure 4). For all sleep outcomes, the 95% CI for the period-adjusted mean difference between exercise and control trials overlapped zero (all main effects of trial $p \geq .545$) and standardized effect sizes were trivial apart from the small effect size for sleep latency ($d = 0.35$). Participant-by-trial interactions were identified for sleep latency and sleep efficiency (both $p \leq .007$) but not total sleep time or actual wake time (both $p \geq .190$) (Table 2). All SD_{IR} estimates were relatively small with a negative SD_{IR} identified for actual wake time, sleep latency, and sleep efficiency (based on estimate 2) and inconsistency in the direction of SD_{IR} was evident between estimates 1 and 2 for actual wake time and sleep efficiency (Table 2). This is consistent with the τ statistic from the random-effects meta-analysis showing that the estimated between-participant variability was negligible (Figure 5). The proportion of participants exhibiting the mean control-adjusted exercise response beyond the MCID was 22% above and 33% below for total sleep time, 11% above and 6% below for actual wake time, 6% above and 0% below for sleep latency, and 22% above and 22% below for sleep efficiency (Figure 5).

Outliers were identified for five trials ($n = 4$ participants) for total sleep time, four trials ($n = 3$ participants) for actual wake time, five trials ($n = 4$ participants) for sleep latency, and four trials ($n = 3$ participants) for sleep efficiency. Interpretation of the Pearson's correlation coefficients between the two replicates of control-adjusted exercise responses after outlier removal was similar for total sleep time ($r = 0.12$ [95% CI = −0.46 to 0.63], $p = .699$), actual wake time ($r = 0.17$ [95% CI = −0.42 to 0.66], $p = .575$), sleep latency ($r = -0.05$ [95% CI = −0.59 to 0.51], $p = .863$), and sleep efficiency ($r = 0.27$ [95% CI = −0.31 to 0.70], $p = .356$). The exclusion of these data did not alter the significance of the participant-by-trial interactions (all $p \geq .115$), and the SD_{IR} for estimates 1 and 2, and the τ statistic from the random-effects meta-analysis (τ range 4–47 minutes) remained similar (Supplementary Table 1).

Comparison 3: average post-intervention (nights 3–5) minus average pre-intervention (nights 1 and 2).

Pearson's correlation coefficients between the two replicates of control adjusted exercise responses were not statistically significant and ranged in magnitude from trivial (total sleep time: $r = -0.04$ [95% CI = −0.50 to 0.43], $p = .867$; sleep latency: $r = 0.07$ [95% CI: −0.41 to 0.52], $p = .784$) to small-to-moderate (actual wake time: $r = 0.41$ [95% CI = −0.07 to 0.74], $p = .090$; sleep efficiency: $r = 0.27$ [95% CI = −0.23 to 0.65], $p = .281$) (Figure 6). For all sleep outcomes, the 95% CIs for the period-adjusted mean difference between exercise and control trials overlapped zero (all main effects of trial $p \geq .322$) and standardized effect sizes were trivial except for the small effect size for total sleep time ($d = 0.27$). Participant-by-trial interactions were not identified

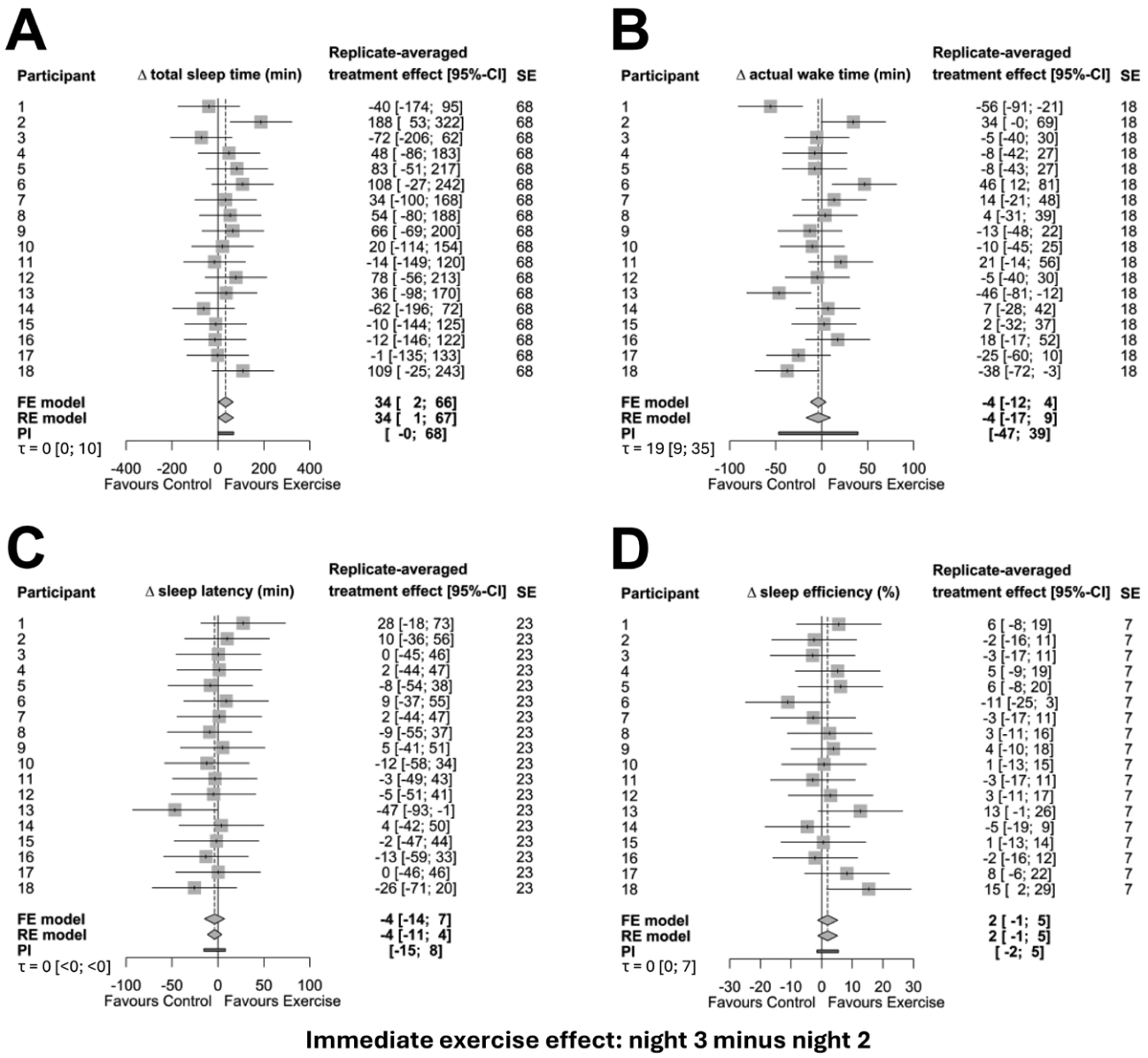


Figure 3. Forest plot displaying the individual participant replicate-averaged exercise effect (night 3 minus night 2) from the random-effects meta-analysis for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 18 men (72 observations/trials). CI, confidence interval; SE, standard error; FE, fixed effect; RE, random effects; PI, prediction interval; τ, tau statistic describing between-participant variability.

for any sleep outcome (all $p \geq .412$) (Table 2). The SD_{IR} for estimate 1 and estimate 2 was small with wide 95% CIs for all outcomes, and the SD_{IR} (estimate 2) for actual wake time and sleep efficiency was negative (Table 2). The τ statistic from the random-effects meta-analysis was close to zero for total sleep time, sleep latency, and sleep efficiency but was higher for actual wake time (τ statistic [95% CI]: 17 [8, 30] minutes) (Figure 7). Most participants exhibited mean control-adjusted exercise responses that did not exceed the MCID (total sleep time: 33% above, 11% below; actual wake time: 11% above, 6% below; sleep latency: 6% above, 6% below; and sleep efficiency: 11% above, 17% below) (Figure 7).

Outliers were identified for 4 trials ($n = 1$ participant) for actual wake time, 12 trials ($n = 3$ participants) for sleep latency, and 8 trials ($n = 2$ participants) for sleep efficiency. A sensitivity analysis removing outliers improved the Pearson's correlation coefficient of the control-adjusted exercise response between replicates for actual wake time ($r = 0.55$ [95% CI = 0.09 to 0.81], $p = .023$). The correlation coefficients after outlier removal were moderate but not statistically significant for sleep latency ($r = –0.33$ [95% CI = –0.72 to 0.22], $p = .227$) and sleep efficiency ($r = 0.34$ [95% CI = –0.18 to 0.72], $p = .193$). Interpretation of the participant-by-trial interactions (all $p \geq .353$), the SD_{IR} for estimates 1 and 2, and the estimated between-participant heterogeneity from the random-effects meta-analysis (τ range 0–19 minutes) were unchanged after outlier removal (Supplementary Table 1).

Discussion

This study adopted a replicate crossover design to investigate the consistency of sleep in response to acute exercise bouts performed on repeated occasions and to examine the extent of true interindividual variability in responses. The primary findings from this study demonstrate that the consistency of control-adjusted

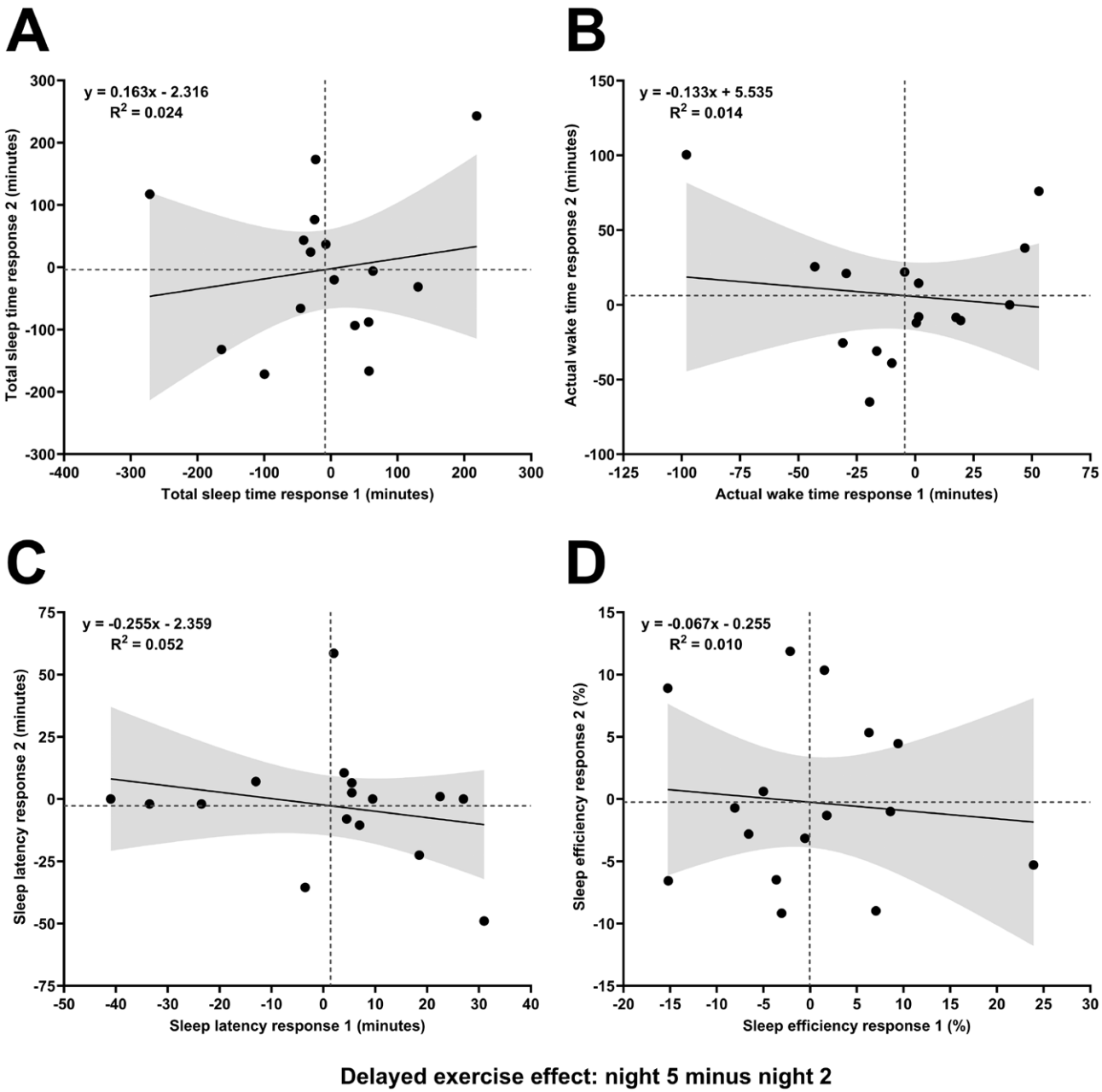


Figure 4. Relationship between exercise and control pre-to-post change scores (night 5 minus night 2) on the two occasions for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 16 men (32 replicates of the control-adjusted exercise response). "Response 1" calculated from the first pair of trials (exercise 1 minus control 1) and "Response 2" calculated from the second pair of trials (exercise 2 minus control 2). Dashed lines depict the mean responses and gray-shaded region represents the 95% CI of the regression line (black solid line).

sleep measured objectively in the home environment to repeated single bouts of exercise is low, and interindividual differences in sleep were small compared with the natural (trial-to-trial) within-subject variability in sleep outcome measures. These findings were evident irrespective of whether sleep was analyzed the night after exercise (immediate), three nights after exercise (delayed), or as an average across the three nights after exercise.

In this study, the mean effects of exercise on total sleep time, actual wake time, sleep latency, and sleep efficiency were relatively low, with 95% CIs that overlapped zero and trivial-to-moderate standardized mean differences. Although beneficial effects of acute exercise have been reported for subjective and objective metrics of sleep quantity and quality [14, 15], the evidence is inconsistent, and the magnitude of response is typically trivial-to-small (Cohen's d = 0.17–0.38) [14]. Notably, there is little experimental evidence that evening physical activity or exercise concluding ~15 minutes to 4 hours before bedtime is detrimental to night-time sleep [17–19]. The mean treatment effect observed in this study appears consistent with existing evidence suggesting that the impact of acute exercise on sleep variables is generally modest but, importantly, is unlikely to negatively impact sleep on the subsequent night. Our study recruited young healthy men without any pre-existing sleep complaints, so it is possible that exercise may promote greater acute sleep benefits in individuals with poor sleep or diagnosed sleep disorders that have a greater capacity for improvement [59, 60].

Analysis of the between-replicate correlation coefficients did not indicate any clear relationship between the paired responses

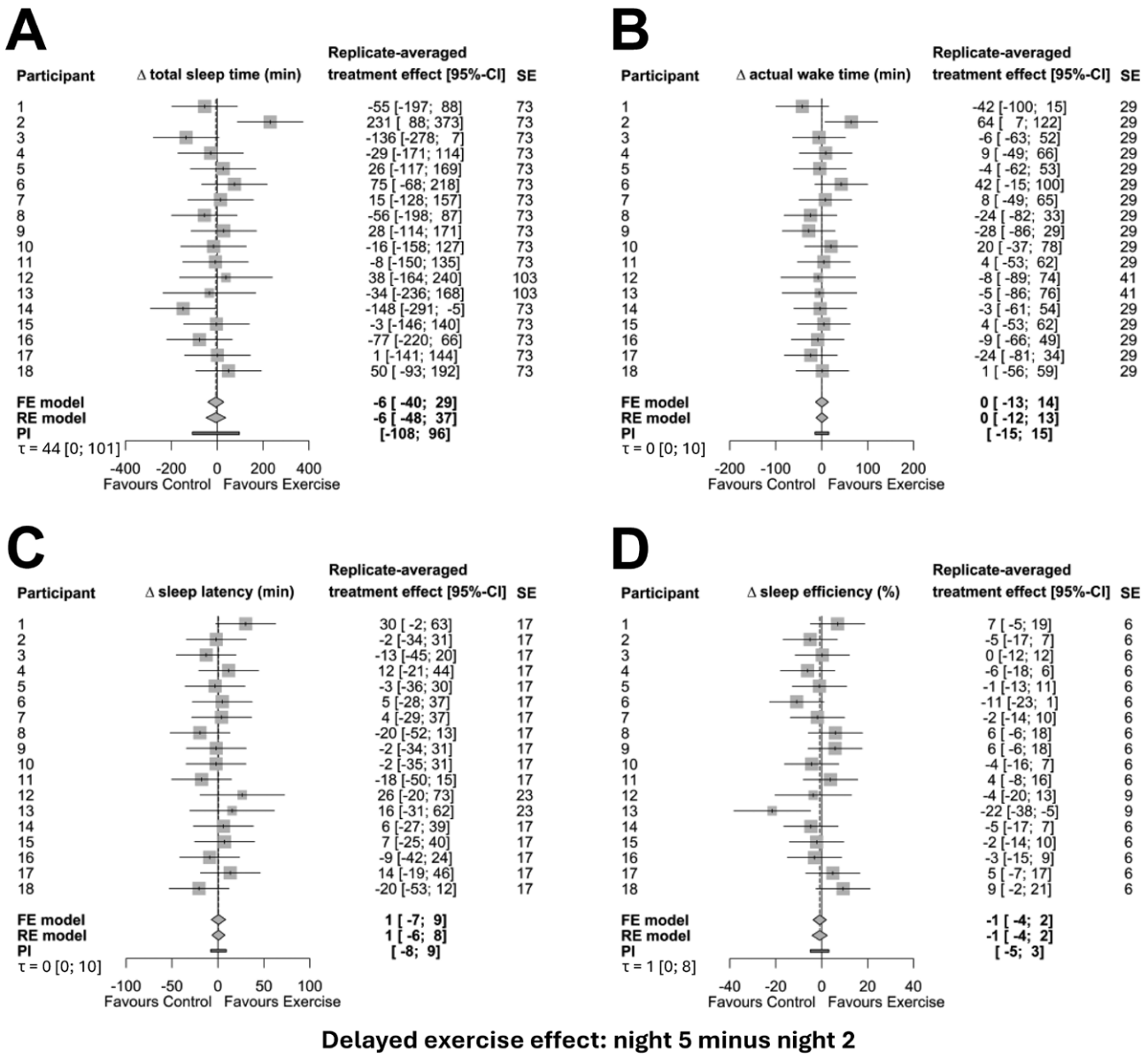


Figure 5. Forest plot displaying the individual participant replicate-averaged exercise effect (night 5 minus night 2) from the random-effects meta-analysis for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 18 men (70 observations/trials). CI, confidence interval; SE, standard error; FE, fixed effect; RE, random effects; PI, prediction interval; τ, tau statistic describing between-participant variability.

for any sleep outcome, suggesting an inconsistency in the effect of exercise on sleep patterns on repeated occasions. This observation was apparent when the immediate, delayed, and average exercise effect was interrogated and persisted after the removal of outliers. Additionally, across the comparisons and outcomes assessed, there were multiple instances where the difference between the replicated control-adjusted exercise responses for a given participant was large and/or in opposing directions on the two occasions. The lack of consistency in sleep in response to acute exercise is likely underscored by measurement error and biological variability resulting from the exercise protocol, free-living sleep assessment, and differences in environmental factors such as circadian variations, diet, and psychological stress [20, 61].

Although no previous studies have investigated the consistency in sleep outcomes after acute exercise, it is well documented that considerable night-to-night variability exists in sleep duration, quality, and timing [32, 33, 35]. Such intra-individual variability appears greater in adults at a younger age possibly due to fluctuating social, working, and family commitments influencing the consistency of sleep–wake cycles [33, 34, 62]. The magnitude of night-to-night variability in wake after sleep onset and number of awakenings appears lower after four months of moderate-intensity walking in older women [63]. Consequently, the extension of the current research design to older adults could be of interest to determine whether meaningful exercise-related response heterogeneity exists in sleep when the level of night-to-night variability in sleep is expected to be lower. Due to the large night-to-night variability in sleep within individuals, it is recognized that single-night-only measures of sleep are unlikely to provide an accurate representation of habitual sleep behavior [32]. Therefore, in addition to focusing on the night immediately after exercise when the influence on sleep was expected to be most pronounced, we also derived a multiple-night average of sleep for the pre-intervention

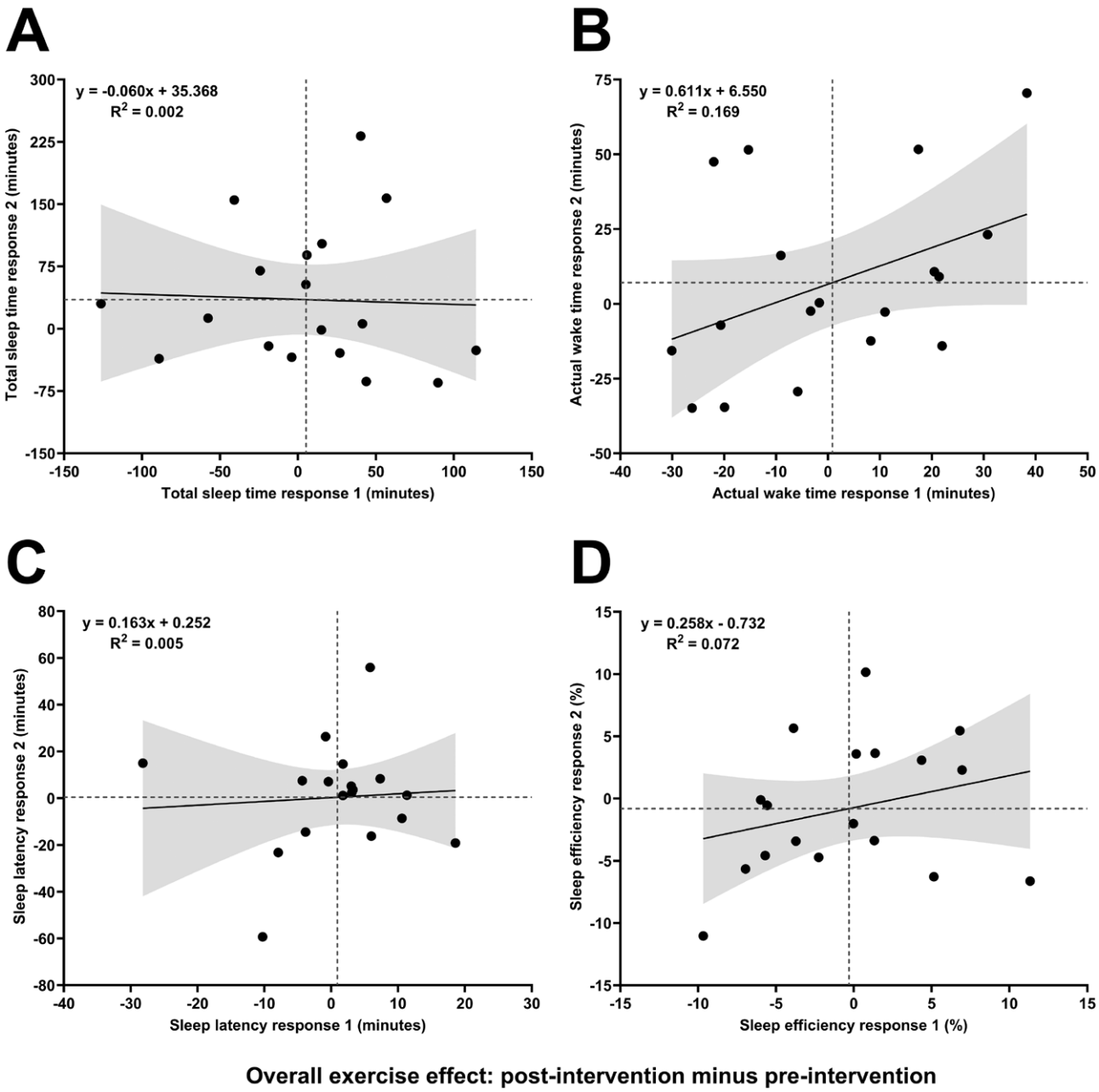


Figure 6. Relationship between exercise and control pre-to-post change scores (average postintervention [nights 3–5] minus average pre-intervention [nights 1 and 2]) on the two occasions for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 18 men (36 replicates of the control-adjusted exercise response). "Response 1" calculated from the first pair of trials (exercise 1 minus control 1) and "Response 2" calculated from the second pair of trials (exercise 2 minus control 2). Dashed lines depict the mean responses and gray-shaded region represents the 95% CI of the regression line (black solid line).

(nights 1 and 2) and post-intervention (nights 3–5) estimates to improve the stability of the sleep assessment.

Another feature of our study design that could contribute to the degree of inherent within-subject variability is whether sleep was assessed on weekdays or weekend days. It was not feasible to control the sleep assessment days across the four trials either within or between participants due to the complexity of the study design, but it is recognized that intra-individual variability in sleep characteristics is typically greater when measured at the weekend than on weekdays [64, 65]. Importantly, sleep data for night 3 (immediate exercise effect) and night 5 (delayed exercise effect) were collected on either all weekday nights or all weekend nights for most participants across the four trials, albeit these nights did consist of a combination of weekday and weekend nights for some participants (≤22%). Arguably, the impact of performing sleep assessments on weekday versus weekend days is likely to be lower in the immediate exercise effect comparison given participants were required to attend the laboratory during the day and follow study protocols until 17:00 on day 2. Furthermore, the inclusion of a multi-night replicate-averaged exercise effect comparison which encompassed a combination of sleep on weekday and weekend nights for all participants, was important to minimize night-to-night variability in the analysis and yielded similar findings to the immediate and delayed

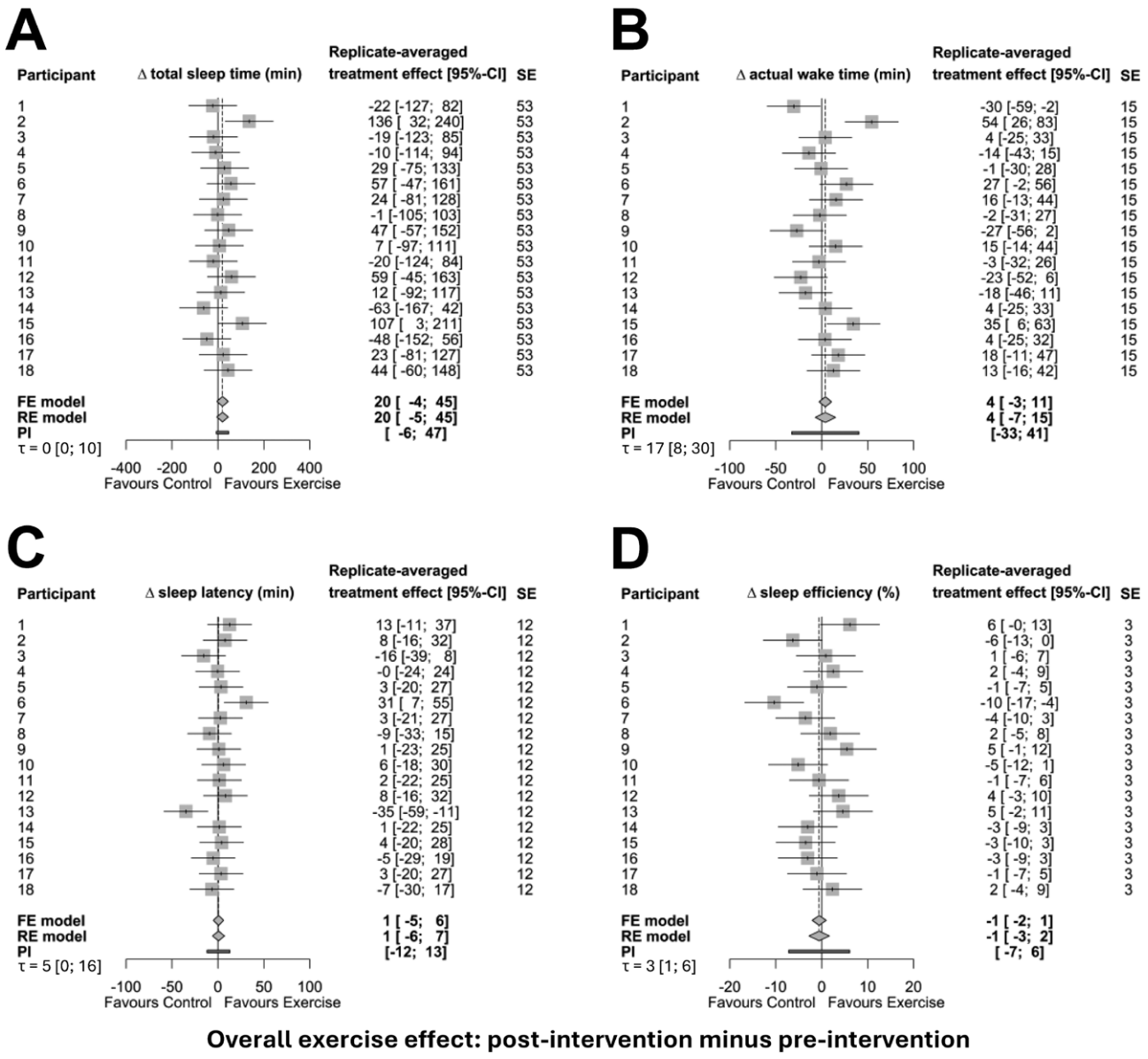


Figure 7. Forest plot displaying the individual participant replicate-averaged exercise effect (post- minus pre-intervention) from the random-effects meta-analysis for: (A) total sleep time; (B) actual wake time; (C) sleep latency; and (D) sleep efficiency in 18 men (72 observations/trials). CI, confidence interval; SE, standard error; FE, fixed effect; RE, random effects; PI, prediction interval; **τ**, tau statistic describing between-participant variability.

exercise effect comparisons. Given the diversity of factors that contribute to night-to-night variability in sleep [34, 62], accounting for natural fluctuations in sleep that occur in the absence of an intervention is paramount when exploring the consistency and magnitude of interindividual differences in response.

Importantly, the adoption of a replicate crossover design permits a formal separation of the participant-by-treatment interaction from random within-participant variability over time [66]. This is the first study to investigate individual differences in sleep in response to acute exercise and therefore contributes novel findings to the literature using a robust study design and analytical approaches [21–24, 43]. Most of the participant-by-trial interactions were trivial, the individual differences *SD*s were relatively small with large uncertainty around the estimates, and the estimated between-participant heterogeneity from the random-effects meta-analysis was negligible for most outcomes. As an exception, a statistically significant participant-by-trial interaction was identified for sleep latency and sleep efficiency for the delayed effect of exercise. However, the **τ** statistic was close to zero, and the individual differences *SD*s for these outcomes, along with several other outcomes (≥50% of outcomes within each comparison), were negative in direction indicating greater variability in the control than exercise trials. It is also notable that the direction of the SD_{IR} calculated from the naïve (estimate 1) and modeling (estimate 2) approaches was inconsistent for several variables, which most likely reflects an imprecision in the estimates due to the degree of measurement error associated with the sleep variables. These findings do not provide any evidence of meaningful interindividual response heterogeneity in sleep in response to acute exercise, which could not be detected due to substantial natural within-subject variability and measurement error over time.

Although this study has several strengths, including the robust study design and statistical approaches and the assessment of

home-based sleep using actigraphy, it is important to highlight some limitations. Our findings may not generalize to other populations including women, older adults, and those with pre-existing sleep conditions. Due to the arduous nature of the study design requiring repeated exercise protocols and multiple outcome assessments, it was not possible to control whether the sleep data were collected on weekday or weekend nights across the trials. Furthermore, participants were instructed to follow their usual sleep routine with no restrictions placed on screen use before sleep, social interactions, or sleep location. After participants left the laboratory on day 2, we were unable to impose restrictions on physical activity engagement or dietary intake which may have influenced the sleep results obtained in the delayed and average comparisons. It is also recognized that wrist-worn actigraphy is a reasonable, albeit imperfect, tool for assessing sleep compared with polysomnography with the low ability for actigraphy devices to detect wakefulness whilst immobile during sleep representing the main limitation [58]. Finally, we did not measure any biological variables implicated in the circadian control of the sleep–wake cycle including melatonin and orexin, which may present a potential direction for future investigations.

In conclusion, the findings of this study suggest that free-living sleep in response to acute exercise was inconsistent when measured on repeated occasions. Considerable trial-to-trial within-subject variability and measurement error prevented the identification of meaningful interindividual response heterogeneity in sleep in response to acute exercise. These findings highlight the importance of researchers accounting for night-to-night variability in sleep before making any inferences on interindividual differences in sleep in response to an intervention.

Supplementary material

Supplementary material is available at *SLEEP Advances* online.

Acknowledgments

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Funding

This research was funded by King Saud bin Abdulaziz University for Health Sciences (Saudi Arabia) and the National Institute for Health and Care Research (NIHR) Leicester Biomedical Research Centre (United Kingdom).

Conflict of interest statement

None.

Data availability

The de-identified data underlying this article will be shared on reasonable request to the corresponding author. The code used for the statistical analysis is available in Supplementary Material.

References

1. Irwin MR. Why sleep is important for health: a psychoneuroimmunology perspective. *Annu Rev Psychol*. 2015;**66**:143–172. doi:10.1146/annurev-psych-010213-115205
2. Hirshkowitz M, Whiton K, Albert SM, *et al*. National Sleep Foundation's updated sleep duration recommendations: final report. *Sleep Health*. 2015;**1**(4):233–243. doi:10.1016/j.sleh.2015.10.004
3. Watson NF, Badr MS, Consensus Conference Panel, *et al*. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *J Clin Sleep Med*. 2015;**11**(8):931–952. doi:10.5664/jcsm.4950
4. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults—United States, 2014. *MMWR Morb Mortal Wkly Rep*. 2016;**65**(6):137–141. doi:10.15585/mmwr.mm6506a1
5. Zhu G, Catt M, Cassidy S, *et al*. Objective sleep assessment in >80,000 UK mid-life adults: associations with sociodemographic characteristics, physical activity and caffeine. *PLoS One*. 2019;**14**(12):e0226220. doi:10.1371/journal.pone.0226220
6. Itani O, Jike M, Watanabe N, Kaneita Y. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. *Sleep Med*. 2017;**32**:246–256. doi:10.1016/j.sleep.2016.08.006
7. Jike M, Itani O, Watanabe N, Buysse DJ, Kaneita Y. Long sleep duration and health outcomes: a systematic review, meta-analysis and meta-regression. *Sleep Med Rev*. 2018;**39**:25–36. doi:10.1016/j.smrv.2017.06.011
8. Liang YY, Ai S, Xue H, *et al*. Joint associations of device-measured sleep duration and efficiency with all-cause and cause-specific mortality: a prospective cohort study of 90 398 UK Biobank participants. *J Gerontol A Biol Sci Med Sci*. 2023;**78**(9):1717–1724. doi:10.1093/gerona/glad108
9. Sletten TL, Weaver MD, Foster RG, *et al*. The importance of sleep regularity: a consensus statement of the National Sleep Foundation sleep timing and variability panel. *Sleep Health*. 2023;**9**(6):801–820. doi:10.1016/j.sleh.2023.07.016
10. Windred DP, Burns AC, Lane JM, *et al*. Sleep regularity is a stronger predictor of mortality risk than sleep duration: a prospective cohort study. *Sleep*. 2024;**47**(1):zsad253. doi:10.1093/sleep/zsad253
11. Buman MP, King AC. Exercise as a treatment to enhance sleep. *Am J Lifestyle Med*. 2010;**4**(6):500–514. doi:10.1177/1559827610375532
12. Huang BH, Inan-Eroglu E, Hamer M, Stamatakis E. Joint associations of device-measured physical activity and sleep duration with cardiometabolic health in the 1970 British Cohort Study. *J Sci Med Sport*. 2020;**23**(12):1191–1196. doi:10.1016/j.jsams.2020.07.012
13. Liang YY, Feng H, Chen Y, *et al*. Joint association of physical activity and sleep duration with risk of all-cause and cause-specific mortality: a population-based cohort study using accelerometry. *Eur J Prev Cardiol*. 2023;**30**(9):832–843. doi:10.1093/eurjpc/zwad060
14. Kredlow MA, Capozzoli MC, Hearon BA, Calkins AW, Otto MW. The effects of physical activity on sleep: a meta-analytic review. *J Behav Med*. 2015;**38**(3):427–449. doi:10.1007/s10865-015-9617-6
15. Kline CE, Hillman CH, Bloodgood Sheppard B, *et al*. Physical activity and sleep: an updated umbrella review of the 2018 Physical Activity Guidelines Advisory Committee report. *Sleep Med Rev*. 2021;**58**:101489. doi:10.1016/j.smrv.2021.101489
16. National Sleep Foundation. *Get moving to get better sleep: healthy sleep habits*. 2022. https://www.thensf.org/get-moving-to-get-better-sleep/. Accessed May 30, 2024.
17. Stutz J, Eiholzer R, Spengler CM. Effects of evening exercise on sleep in healthy participants: a systematic review and

meta-analysis. *Sports Med*. 2019;**49**(2):269–287. doi:10.1007/s40279-018-1015-0
18. Frimpong E, Mograss M, Zvionow T, Dang-Vu TT. The effects of evening high-intensity exercise on sleep in healthy adults: a systematic review and meta-analysis. *Sleep Med Rev*. 2021;**60**:101535. doi:10.1016/j.smrv.2021.101535
19. Kahn M, Korhonen T, Leinonen L, *et al*. Is it time we stop discouraging evening physical activity? New real-world evidence from 150,000 nights. *Front Public Health*. 2021;**9**:772376. doi:10.3389/fpubh.2021.772376
20. Swinton PA, Hemingway BS, Saunders B, Gualano B, Dolan E. A statistical framework to interpret individual response to intervention: paving the way for personalized nutrition and exercise prescription. *Front Nutr*. 2018;**5**:41. doi:10.3389/fnut.2018.00041
21. Atkinson G, Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol*. 2015;**100**(6):577–588. doi:10.1113/EP085070
22. Atkinson G, Williamson P, Batterham AM. Issues in the determination of "responders" and 'non-responders' in physiological research. *Exp Physiol*. 2019;**104**(8):1215–1225. doi:10.1113/EP087712
23. Senn S, Rolfe K, Julious SA. Investigating variability in patient response to treatment--a case study from a replicate cross-over study. *Stat Methods Med Res*. 2011;**20**(6):657–666. doi:10.1177/0962280210379174
24. Senn S. Mastering variation: variance components and personalised medicine. *Stat Med*. 2016;**35**(7):966–977. doi:10.1002/sim.6739
25. Gewandter JS, McDermott MP, He H, *et al*. Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine. *Clin Pharmacol Ther*. 2019;**106**(1):204–210. doi:10.1002/cpt.1372
26. Hopkins WG. Individual responses made easy. *J Appl Physiol (1985)*. 2015;**118**(12):1444–1446. doi:10.1152/japplphysiol.00098.2015
27. Kelley GA, Kelley KS, Stauffer BL. Effects of resistance training on body weight and body composition in older adults: an inter-individual response difference meta-analysis of randomized controlled trials. *Sci Prog*. 2023;**106**(2):368504231179062. doi:10.1177/00368504231179062
28. Goltz FR, Thackray AE, King JA, Dorling JL, Atkinson G, Stensel DJ. Interindividual responses of appetite to acute exercise: a replicated crossover study. *Med Sci Sports Exerc*. 2018;**50**(4):758–768. doi:10.1249/MSS.0000000000001504
29. Goltz FR, Thackray AE, Atkinson G, *et al*. True interindividual variability exists in postprandial appetite responses in healthy men but is not moderated by the FTO genotype. *J Nutr*. 2019;**149**(7):1159–1169. doi:10.1093/jn/nxz062
30. Sundström J, Lind L, Nowrouzi S, *et al*. Heterogeneity in blood pressure response to 4 antihypertensive drugs: a randomized clinical trial. *JAMA*. 2023;**329**(14):1160–1169. doi:10.1001/jama.2023.3322
31. Shen T, Thackray AE, King JA, *et al*. Are there interindividual responses of cardiovascular disease risk markers to acute exercise? A replicate crossover trial. *Med Sci Sports Exerc*. 2024;**56**(1):63–72. doi:10.1249/MSS.0000000000003283
32. Knutson KL, Rathouz PJ, Yan LL, Liu K, Lauderdale DS. Intra-individual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. *Sleep*. 2007;**30**(6):793–796. doi:10.1093/sleep/30.6.793
33. Dillon HR, Lichstein KL, Dautovich ND, Taylor DJ, Riedel BW, Bush AJ. Variability in self-reported normal sleep across the adult age span. *J Gerontol B Psychol Sci Soc Sci*. 2015;**70**(1):46–56. doi:10.1093/geronb/gbu035
34. Bei B, Wiley JF, Trinder J, Manber R. Beyond the mean: a systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. *Sleep Med Rev*. 2016;**28**:108–124. doi:10.1016/j.smrv.2015.06.003
35. Chouraki A, Tournant J, Arnal P, Pépin JL, Bailly S. Objective multi-night sleep monitoring at home: variability of sleep parameters between nights and implications for the reliability of sleep assessment in clinical trials. *Sleep*. 2023;**46**(5):zsac319. doi:10.1093/sleep/zsac319
36. Couillard C, Bergeron N, Prud'homme D, *et al*. Gender difference in postprandial lipemia: importance of visceral adipose tissue accumulation. *Arterioscler Thromb Vasc Biol*. 1999;**19**(10):2448–2455. doi:10.1161/01.atv.19.10.2448
37. Gill JM, Malkova D, Hardman AE. Reproducibility of an oral fat tolerance test is influenced by phase of menstrual cycle. *Horm Metab Res*. 2005;**37**(5):336–341. doi:10.1055/s-2005-861481
38. Dwan K, Li T, Altman DG, Elbourne D. CONSORT: 2010 statement: extension to randomised crossover trials. *BMJ*. 2019;**366**:l4378. doi:10.1136/bmj.l4378
39. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;**14**(5):377–381.
40. Frayn KN. Calculation of substrate oxidation rates in vivo from gaseous exchange. *J Appl Physiol Respir Environ Exerc Physiol*. 1983;**55**(2):628–634. doi:10.1152/jappl.1983.55.2.628
41. Oakley NR. Validation with polysomnography of the Sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system. Technical Report to Mini Mitter Co., Inc. 1997. https://www.camntech.com/Engineering/MotionWare/AW95001.pdf
42. Rupp TL, Balkin TJ. Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults. *Behav Res Methods*. 2011;**43**(4):1152–1160. doi:10.3758/s13428-011-0098-4
43. Senn S. The analysis of continuous data from n-of-1 trials using paired cycles: a simple tutorial. *Trials*. 2024;**25**(1):128. doi:10.1186/s13063-024-07964-7
44. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
45. Higgins JPT, Li T, Deeks JJ, eds. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, *et al*., eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated August 2023)*. Cochrane: London; 2023. www.training.cochrane.org/handbook
46. SAS Institute. SAS® OnDemand for Academics; 2023. [cited June 12, 2023]. https://www.sas.com/en_us/software/on-demand-for-academics.html
47. Schabenberger O. Mixed model influence diagnostics. In: Proceedings of the twenty ninth annual SAS Users Group International Conference; May 9–12, 2004. Cary, NC: SAS Institute; 2004: 189–129.
48. West BT, Galecki AT. An overview of current software procedures for fitting linear mixed models. *Am Stat*. 2012;**65**(4):274–282. doi:10.1198/tas.2011.11077
49. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;**14**:25. doi:10.1186/1471-2288-14-25
50. Schwarzer G. An R package for meta-analysis. *R News*. 2007;**40**:45.
51. Veroniki AA, Jackson D, Viechtbauer W, *et al*. Methods to estimate the between-study variance and its uncertainty in

meta-analysis. *Res Synth Methods*. 2016;**7**(1):55–79. doi:10.1002/jrsm.1164
52. Langan D, Higgins JPT, Jackson D, *et al*. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods*. 2019;**10**(1):83–98. doi:10.1002/jrsm.1316
53. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med*. 2007;**26**(1):37–52. doi:10.1002/sim.2514
54. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;**1**(2):97–111. doi:10.1002/jrsm.12
55. Borenstein M. Avoiding common mistakes in meta-analysis: Understanding the distinct roles of Q, I-squared, tau-squared, and the prediction interval in reporting heterogeneity. *Res Synth Methods*. 2024;**15**(2):354–368. doi:10.1002/jrsm.1678
56. Cook JA, Julious SA, Sones W, *et al*. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Trials*. 2018;**19**(1):606. doi:10.1186/s13063-018-2884-0
57. Smith MT, McCrae CS, Cheung J, *et al*. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med*. 2018;**14**(7):1209–1230. doi:10.5664/jcsm.7228
58. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. *Med Sci Sports Exerc*. 2019;**51**(7):1538–1557. doi:10.1249/MSS.0000000000001947
59. Youngstedt SD. Ceiling and floor effects in sleep research. *Sleep Med Rev*. 2003;**7**(4):351–365. doi:10.1053/smrv.2001.0239
60. Passos GS, Poyares D, Santana MG, Garbuio SA, Tufik S, Mello MT. Effect of acute physical exercise on patients with chronic primary insomnia. *J Clin Sleep Med*. 2010;**6**(3):270–275.
61. Chrzanowski-Smith OJ, Piatrikova E, Betts JA, Williams S, Gonzalez JT. Variability in exercise physiology: can capturing intra-individual variation help better understand true inter-individual responses? *Eur J Sport Sci*. 2020;**20**(4):452–460. doi:10.1080/17461391.2019.1655100
62. Messman BA, Wiley JF, Yap Y, *et al*. How much does sleep vary from night-to-night? A quantitative summary of intraindividual variability in sleep by age, gender, and racial/ethnic identity across eight-pooled datasets. *J Sleep Res*. 2022;**31**(6):e13680. doi:10.1111/jsr.13680
63. Breneman CB, Kline CE, West DS, *et al*. The effect of moderate-intensity exercise on nightly variability in objectively measured sleep parameters among older women. *Behav Sleep Med*. 2019;**17**(4):459–469. doi:10.1080/15402002.2017.1395337
64. Tworoger SS, Davis S, Vitiello MV, Lentz MJ, McTiernan A. Factors associated with objective (actigraphic) and subjective sleep quality in young adult women. *J Psychosom Res*. 2005;**59**(1):11–19. doi:10.1016/j.jpsychores.2005.03.008
65. Aili K, Åström-Paulsson S, Stoetzer U, Svartengren M, Hillert L. Reliability of actigraphy and subjective sleep measurements in adults: the design of sleep assessments. *J Clin Sleep Med*. 2017;**13**(1):39–47. doi:10.5664/jcsm.6384
66. Senn S. Sample size considerations for n-of-1 trials. *Stat Methods Med Res*. 2019;**28**(2):372–383. doi:10.1177/0962280217726801