nature portfolio

Peer Review File

Dose-response relationship between evening exercise and sleep

Corresponding Author: Dr Josh Leota

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This was an interesting study that leveraged Whoop band data over a 1-year period to examine the association between exercise timing and intensity on measures of sleep onset, duration, and quality as well as nocturnal autonomic activity. The paper was well written and the results clearly conveyed. I believe the findings make a significant contribution to what is known about the topic. I have the following comments/questions for the authors.

• Lines 211-212. Why was a minimum of 50 exercise sessions required for inclusion?

• Exercise Timing Assessment (lines 223-231) – Was there any way to estimate the time a participant went to bed with the intention of sleep for a given day?

• Lines 233-245. I'm unfamiliar with the method used to measure exercise strain. How were the cutpoints for the different strain categories selected? The category descriptions don't seem to reflect typical PA categories. For example, the description of the moderate category in Table 2 (45 min at 85% (zone 4) of HRmax) would be equivalent to half the weekly amount of PA required to meet US PA guidelines. I suspect that the vast majority of US adult population would be performing exercise in the light PA category.

• Lines 104-110. The finding that HRV was lower with higher strains even when exercise was performed >6 hr before bed was surprising to me. On lines 133-134, the authors rightly suggest that "To avoid disruptions to sleep, individuals should allow for sufficient recovery time proportionate to the level of strain achieved in an evening exercise session." However, given the lower HRV observed for higher intensity exercise up to 10 hrs prior to bed, what would be a sufficient duration of recovery to avoid this HRV depression?
Note: While beyond the scope of this paper, it would be interesting to see if the suppressed HRV accumulates across successive days of high strain exercise.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Overall, I am pleased to see such a well-conducted analysis based on a large sample, where individuals' sleep and evening exercise were tracked for one year. Below are some points that may help improve the presentation and interpretation of your study results:

Does habitual sleep onset refer to the average across the year? Would this approach be appropriate, considering seasonal variations in sleep/wake cycles?

Did you include the week following daylight saving time clock changes in your analysis? A recent study published in the Journal of Sleep Research suggests that catch-up sleep occurs after both transitions, with sleep duration slightly higher from Monday to Friday compared to before, although this effect was less pronounced in autumn (https://onlinelibrary.wiley.com/doi/10.1111/jsr.14335).

Are the observed associations between evening exercise timing and sleep outcomes sex-specific? Please include relevant figures in the supplement.

Are the observed associations between evening exercise timing and sleep outcomes age-specific? Consider merging the middle and upper age range groups, as the latter is too small. Again, please include relevant figures in the supplement.

Given the seasonal effects on sleep, I would like to see whether the observed associations between evening exercise timing and sleep outcomes vary by season. For example, evening exercise may be more likely to occur outdoors in summer, when temperatures are higher and evenings are brighter.

Morning and early afternoon exercise, even based on smaller samples, has been proposed to advance the onset of sleep. Therefore, I wonder whether morning exercise (for which you have data, correct?) could at least partially offset some of the observed detrimental effects of evening exercise on sleep outcomes.

Please add a table to the supplement comparing basic characteristics between those included in the study and those excluded.

Exercise, particularly high-intensity or maximal activities, may be more physiologically exhausting for individuals with a high BMI (e.g., ≥30 kg/m²) compared to those with normal weight. Thus, I would like to see whether the observed associations between evening exercise and sleep outcomes vary by BMI status.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have done a stellar job in responding to my previous comments and suggestions. I also found the additional supplemental analyses performed in response to both reviewers comments to be informative. Excellent work!

Reviewer #2

(Remarks to the Author)
Dear Author Team,

It has been a pleasure reading the revised version of your manuscript. You have addressed all the points raised by me (and, as far as I can see, also those raised by the other reviewer) in a professional and satisfactory manner.

That being said, I have no further comments and hope your paper receives the attention it deserves, should it be accepted for publication.

Point-by-point response to reviewers' comments

Manuscript Number: NCOMMS-24-65275

Reviewer 1:

This was an interesting study that leveraged Whoop band data over a 1-year period to examine the association between exercise timing and intensity on measures of sleep onset, duration, and quality as well as nocturnal autonomic activity. The paper was well written and the results clearly conveyed. I believe the findings make a significant contribution to what is known about the topic.

We are pleased Reviewer 1 had such a positive response to our work, and we are gratified they considered our study interesting, our manuscript clear and well written, and our findings of significant value. We also appreciate their insightful questions and suggestions below.

R1.1. Why was a minimum of 50 exercise sessions required for inclusion?

We chose this inclusion criterion for two reasons. First, each participant should have a sufficient number of exercise data points to allow for robust within-person comparisons of different timings and doses (i.e., levels of strain) of exercise. A recent systematic review of daily associations between sleep and physical activity suggested 50 days of observations are required to obtain robust within-person (individual level) results (Atoui et al., 2021).

Second, we wanted our sample to reflect physically active individuals who exercised approximately once per week on average. Physically active individuals are more likely to engage in strenuous exercise compared to inactive individuals. Further, physically active individuals have developed a baseline tolerance for the physiological consequences of exercise *per se*, such that we can more effectively isolate the relationship between *strenuous evening* exercise and sleep and nocturnal autonomic function. Consider an individual who only exercises a few times per year, with one of these sessions occurring in the late evening. It would be difficult to disentangle the physiological response to an individual's first exercise session in a long time from the specific effects related to the proximity of that exercise session to the individual's bedtime.

This rationale is now included in the Methods, page 13:

Eligibility criteria required participants be ≥18 years of age, wear the biometric device for a minimum of 350 days (95.6% of total possible wear-time), and register ≥50 exercise activities within the analysis window of 10-h before to 2-h after habitual sleep onset to construct a sample of physically active individuals (i.e., with an exercise frequency of approximately one or more activities weekly) and to allow for within-person comparisons of different exercise timings and levels of exercise strain.

The Abstract, Introduction, and Discussion also now refer to "physically active" individuals.

Abstract, page 2:

This study examines the association between strenuous evening exercise and objective sleep, using data from 14,689 physically active individuals who wore a biometric device for 365 days (4,084,354 person-nights).

Introduction, page 3:

Here, in a large multi-national sample of physically active adults, we examined the influence of exercise strain on the association between exercise timing and objectively measured sleep and nocturnal autonomic activity using a validated wrist-worn biometric device in free-living conditions.

Discussion, page 12:

This study examined data from a large sample (14,689 physically active participants) across one year (4,084,354 person-nights) in a real-world setting.

R1.2. Was there any way to estimate the time a participant went to bed with the intention of sleep for a given day?

We could not estimate the time participants intended to sleep. This limitation is not unique to the WHOOP device, nor wearable devices in general, as this limitation is present in any epidemiological methodology that does not require a person to subjectively report, manually input, or otherwise be observed when sleep is first attempted. Given our large sample, long study period, and free-living conditions, it was not feasible to require participants fill out daily logs of when they intended to sleep. Even if it were feasible, this requirement could have itself caused participants to modify their behavior (i.e., the Hawthorne effect) or biased the final sample. Our intention was to collect data as passively as possible so we could observe participant behavior in naturalistic settings.

We acknowledge that the time one intends to sleep and sleep onset latency are important sleep characteristics to consider, especially as strenuous evening exercise may heighten presleep cognitive hyperarousal and delay sleep initiation. We now recommend, in the Discussion, that future field and laboratory studies could build on our research by using measurement methods that either include manual adjustment of 'lights out' time, or use behavioral devices that record behavioral responses to auditory stimuli or the depressing of a microswitch (reviewed in Scott et al., 2020).

In the Discussion, page 11:

The present study had limitations that should be considered. The biometric device did not detect when sleep was first attempted; therefore, we were unable to estimate the time participants began to attempt sleep or calculate sleep onset latency. Future research could collect self-reported timing of attempted sleep or use behavioural devices that record conscious responses to stimuli to investigate whether heightened

sympathetic arousal associated with strenuous evening exercise increases sleep onset latency.

R1.3.1. I'm unfamiliar with the method used to measure exercise strain. How were the cutpoints for the different strain categories selected? The category descriptions don't seem to reflect typical PA categories. For example, the description of the moderate category in Table 2 (45 min at 85% (zone 4) of HRmax) would be equivalent to half the weekly amount of PA required to meet US PA guidelines.

Summated heart rate zone scores (SHRZS; first proposed by Edwards, 1993) were used to quantify exercise strain (i.e., cardiometabolic load). This calculation method, which has been used in previous studies to quantify cardiovascular exercise strain (e.g., Borresen & Lambert, 2008; Borresen & Lambert, 2009), considers both the intensity (heart-rate zone) and duration of exercise and is therefore distinct from traditional physical activity metrics that typically estimate exercise intensity alone. This distinction is important for testing our central hypothesis, as the cardiometabolic load and associated physiological response is higher after a 90-minute high-intensity session compared to a 10-minute high-intensity session. The heart-rate zones are calculated as a percentage of each individual's maximal heart rate and therefore have the added advantage of reflecting *relative* cardiometabolic load as opposed to absolute load (i.e., two individuals may respond to the same prescribed workout differently). The SHRZS method also allows for a more accurate quantification of exercises that involve various exercise intensities (e.g., ball sports, interval training).

The cut points used to classify SHRZS into exercise strain groups (Light, Moderate, High, Maximal) are designed to reflect common groups of exercises that vary in cardiometabolic load and that based on our hypothesis would be expected to impact subsequent sleep differently. The Light strain group reflects exercises that produce minimal stress on the body and allow for active recovery (e.g., a 30-minute brisk walk). The Moderate strain group reflects exercises that produce moderate stress on the body and aid fitness maintenance (e.g., 45-minute group fitness class). The High strain group reflects exercises that produce significant stress on the body with the intention of building fitness gains (e.g., 10-mile run at 85% of maximal heart rate). Finally, the Maximal strain group is reserved only for exercise bouts that produce such high levels of stress on the body that it may be challenging to recover from in the short-term (e.g., competitive hockey match, half marathon).

The absolute values of the cut off points are not intended to be discrete boundaries. Rather, their primary purpose is to allow for classification of a heterogenous set of activities into interpretable relative exercise strain groups based on cardiometabolic load. Importantly, we observe a clear *dose-dependent* relationship with exercise strain group and all outcomes. This is most evident with the cardiovascular measures: as exercise strain groups increase in magnitude, nocturnal resting heart rate increases and nocturnal HRV decreases regardless of exercise timing. This relationship strongly suggests the magnitude of exercise strain groups reflects the magnitude of physical stress produced by the exercise.

To test whether our findings are robust to a different method of quantification of exercise strain, we repeated the primary analyses using the proprietary WHOOP strain score and used WHOOP's own cut off points to classify exercise strain. Results were extremely similar to the results reported in the manuscript (Figure 1 below). The small differences between these results are partially attributable to our inability to combine algorithmically derived WHOOP strain scores from multiple exercises that were logged in quick succession; another advantage of using SHRZS.

Finally, the SHRZS method also has an open-source and straightforward formula that can be replicated in future studies using any device that accurately records heart rate. For these reasons, and consistent with Open Science guidelines and in the interest of reproducibility, we believe that the SHRZS method is appropriate.

We now include this information in the Methods, page 15:

Exercise strain was quantified using summated-heart-rate-zones scores (SHRZS). The duration (minutes) spent in each heart rate zone was multiplied by a corresponding factor for each zone (i.e., zone 1=1; zone 2=2; etc.) and then summated. This method has an open-source formula that can be replicated in future studies using any device that accurately records heart rate. SHRZS were then categorized into four groups designed to reflect common groups of exercises that vary in cardiometabolic load and that based on our hypothesis would be expected to impact subsequent sleep differently: <116=*light*; 116 to <214=*moderate*; 214 to 461=*high*; >461=*maximal* (Table 2).

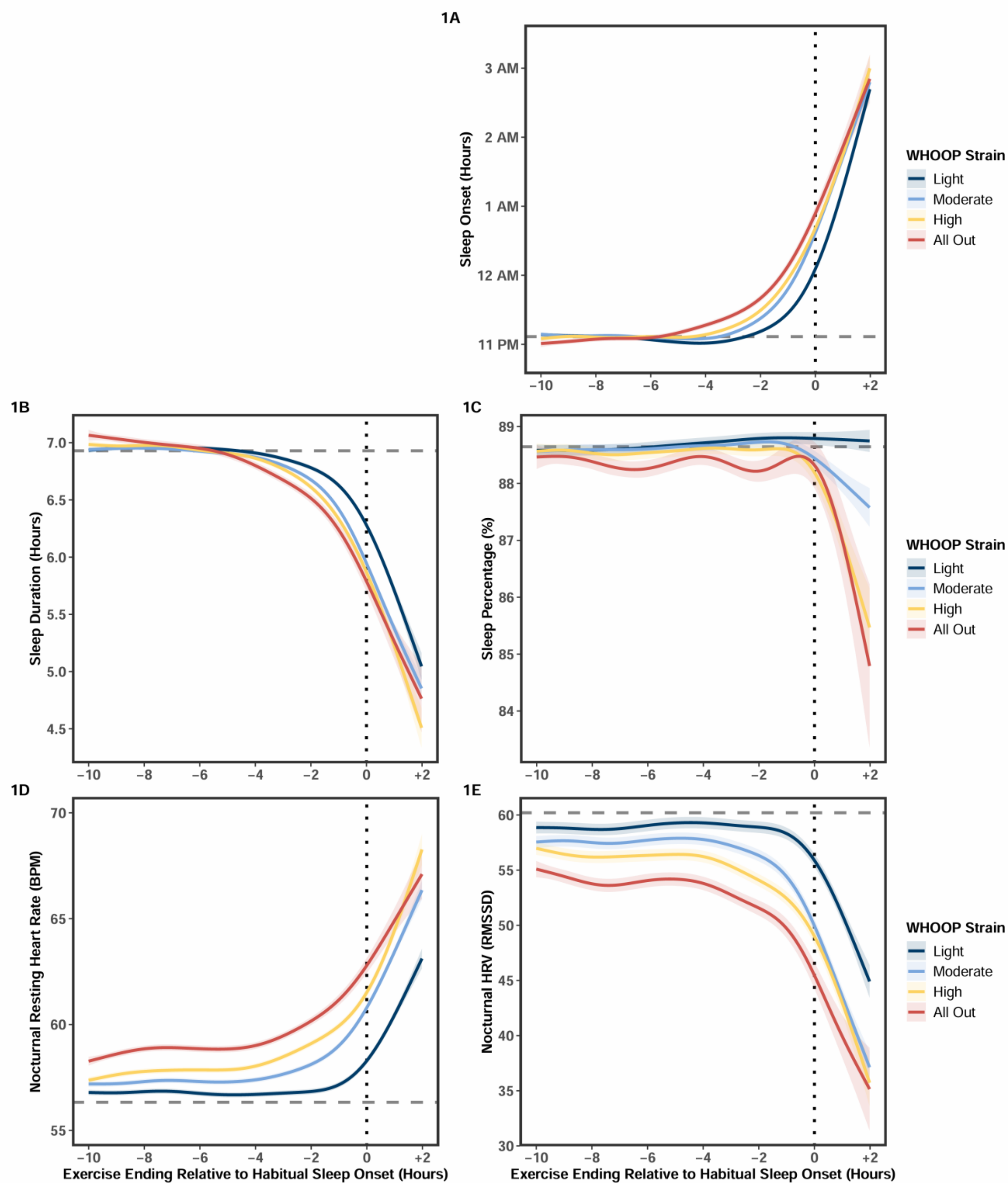

Figure 1. GAMMs demonstrating the relationship between exercise timing and **A.** sleep onset, **B.** sleep duration (in hours), **C.** sleep quality, **D.** nocturnal resting heart rate, and **E.** nocturnal HRV at different levels of exercise strain **derived from the proprietary WHOOP Strain Score** (Note WHOOP, Inc. label the highest strain group "All Out"). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dotted line indicates the mean outcome variable on non-exercise days. Vertical black dotted line indicates habitual sleep onset.

R1.3.2. I suspect that the vast majority of US adult population would be performing exercise in the light PA category.

We agree that most exercises performed by the majority of the US adult population would likely fall into the Light exercise strain category. This is consistent with previous studies (reviewed in Stutz et al., 2019; Frimpong et al., 2021) reporting no relationship between evening exercise and sleep. We estimated the exercise strain group for the 44 unique exercise conditions from the 29 unique studies reviewed in Stutz et al. (2019) and Frimpong et al. (2021) and found that 82% of the exercises would be categorised as Light or Moderate strain in the current study (See Strain_Review_NC_R1 Excel file). Taken together, prior research and our findings suggest most adults should be able to continue to engage in lighter strain exercise up to ~2-h before habitual sleep onset without experiencing sleep disruption.

However, one key strength of our study is that we captured 317,896 instances of High or Maximal strain exercise; a number that far exceeds any previous study. These data are critical to testing our novel hypothesis and the primary contribution of our work that the combination of higher exercise strain and later exercise timing dose-dependently predicts disruptions to sleep and nocturnal autonomic activity.

We now address this point in the Discussion, page 12:

Finally, selection of only physically active adults created a sample that is not representative across the spectrum of physical activity levels. Future research could explore the extent to which our findings apply to the lower end of the physical activity spectrum. We predict the relationship between strenuous evening exercise and sleep to be even stronger in this group, as post-exercise autonomic recovery typically occurs less rapidly in individuals who train less frequently,[21] have lower levels of cardiovascular fitness,[27] and have lower resting vagal control.[33]

R1.4.1. The finding that HRV was lower with higher strains even when exercise was performed >6 hr before bed was surprising to me. On lines 133-134, the authors rightly suggest that "To avoid disruptions to sleep, individuals should allow for sufficient recovery time proportionate to the level of strain achieved in an evening exercise session." However, given the lower HRV observed for higher intensity exercise up to 10 hrs prior to bed, what would be a sufficient duration of recovery to avoid this HRV depression?

This is an excellent point, and we acknowledge we could have been more precise in our original statement. Prior research suggests the time required for *complete* autonomic recovery from a single exercise session ranges from 24–48 hours depending on the level of exercise strain (reviewed in Stanley et al., 2013). This process is gradual and progressive and concludes when cardiac parasympathetic activity (often approximated by measuring HRV) returns to, or rebounds above, pre-exercise levels (Buchheit et al., 2009). Importantly, HRV depression does not necessarily indicate sleep disruption is likely, especially when outside of the acute recovery period (0–4 hours post-exercise). For example, one study, which prescribed exercise in the early evening (~4-h before bed) found increased exercise intensity and duration impacted cardiac autonomic modulation (e.g., elevated heart rate and supressed HRV) but *did not* affect sleep quality (Myllymäki et al., 2012). Our results were similar; when completed at least four hours before bed, higher strain exercise was dose-dependently associated with higher nocturnal resting heart rate and lower nocturnal HRV but was only minimally—if at all—associated with changes to sleep (Figures 1 and 2 of the manuscript). Therefore, to answer the question, "*given the lower HRV observed for higher intensity exercise up to 10 hrs prior to bed, what would be a sufficient duration of recovery to avoid this HRV depression*", it may be that there is no sufficient duration of recovery within a single day to avoid *some* level of HRV depression from certain higher strain exercise bouts, but—according to our data—there may be an adequate duration to minimise *sleep* disruption.

To verify this, **we reanalysed the cardiac data but included exercises concluding 10–16 hours before habitual sleep onset** (e.g., concluding between 7:00 AM and 12:00 PM for individuals with an 11:00 PM average sleep onset; Figure 2 below). Small changes to nocturnal resting heart rate and HRV were still evident many hours after cessation of high strain exercise, suggesting *complete* autonomic recovery may not have been achieved.

In response to this comment, we have revised the language in the manuscript as follows:

In the Discussion, page 10:

To <u>minimise</u> disruptions to sleep, individuals should allow for <u>adequate</u> recovery time proportionate to the level of strain achieved in an evening exercise session.

In the Conclusion, page 12:

To optimize sleep, individuals should aim to complete bouts of exercise four or more hours before their sleep onset to <u>minimise</u> potential adverse consequences to their sleep.

We have also added a paragraph discussing the HRV findings, page 10:

When exercise concluded 6–10 hours before habitual sleep onset, nocturnal HRV was slightly lower compared to non-exercise days. These differences were strain-dependent and minimal (e.g., a 1.56–4.52 unit decrease in HRV), suggesting healthy nocturnal autonomic functioning under these conditions. Nonetheless, this finding aligns with prior research demonstrating that the time required for *complete* autonomic recovery from a single exercise session can exceed 24 hours, especially following strenuous exercise.[27]

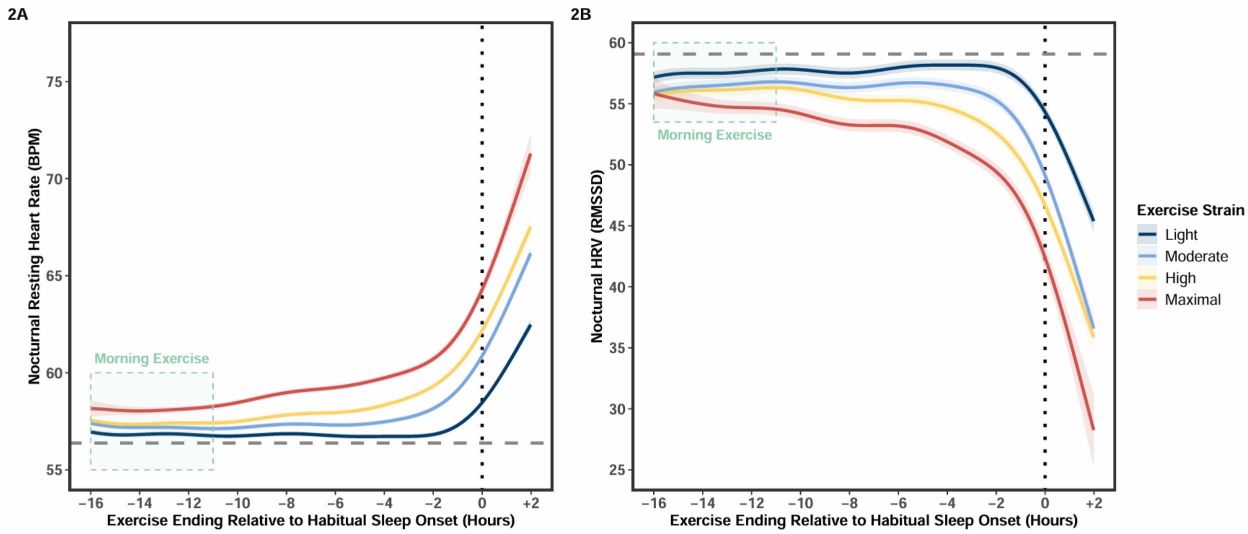

Figure 2. GAMMs demonstrating the relationship between exercise timing and nocturnal **A.** resting heart rate and **E.** HRV at different levels of exercise strain. **Morning exercises (i.e., 10–16 hours before habitual sleep onset) highlighted in teal**. Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dotted line indicates the mean outcome variable on non-exercise days. Vertical black dotted line indicates habitual sleep onset.

R1.4.2. Note: While beyond the scope of this paper, it would be interesting to see if the suppressed HRV accumulates across successive days of high strain exercise.

We agree this would be beyond the scope of this article, but an excellent suggestion for future research to consider. We now include this suggestion in the Discussion, page 10:

Future research should examine how continuous measures of HRV recovery following exercise relate to subsequent sleep, and whether this relationship is moderated by successive days of high strain exercise.

Reviewer 2:

Overall, I am pleased to see such a well-conducted analysis based on a large sample, where individuals' sleep and evening exercise were tracked for one year. Below are some points that may help improve the presentation and interpretation of your study results.

We are grateful for Reviewer 2's suggestions and appreciate their positive response to our research.

R2.1. Does habitual sleep onset refer to the average across the year? Would this approach be appropriate, considering seasonal variations in sleep/wake cycles?

This is an excellent question. We agree that a yearly average would not account for variation in sleep/wake cycles (which varies seasonally and on weekdays versus weekends), which is why we defined habitual sleep onset timing as participant's average sleep onset time of the corresponding calendar month, and the corresponding day of the week (i.e., weekday night or weekend night).

For example, if a participant finished their exercise bout at 5:00 PM on a Tuesday in April, the exercise timing value would be the time, in hours, from 5:00 PM until the participant's average sleep onset time of all weekdays in April. This method accounts for both seasonal variations in sleep/wake cycles as well as social jet lag (i.e., the discrepancy in sleep timing between work and free days). Note that results do not significantly change when habitual sleep onset timing is set to the participant's mean sleep onset time of the corresponding quarter (3-month window).

We outline this definition of habitual sleep onset timing in the Methods, page 14:

Exercise timing was operationalised using two reference points. In the primary analysis, exercise timing represented the conclusion of exercise relative to the participant's *habitual* sleep onset. Individuals display a high degree of heterogeneity in sleep/wake timing. What constitutes an *evening* exercise will depend on the typical sleep/wake patterns of the individual. Further, an individual's sleep/wake timing can vary by season and by weekday/weekend. Therefore, *habitual* sleep onset was quantified as the participant's mean sleep onset of the corresponding calendar month, and whether it was a weekday or weekend to account for seasonal variation and social jet lag. In the secondary analysis, exercise timing represented the conclusion of exercise relative to the participant's sleep onset *on that night*. This method provides important context for how long it can take to recover from varying levels of exercise strain before sleep.

Given the Methods appear at the end of the manuscript, we also define habitual sleep onset timing at first use, in the Results, page 4:

When exercise concluded ≥6-h before social jet lag- and seasonality-adjusted habitual sleep onset (i.e., the average sleep onset corresponding to weekdays or weekends, as

appropriate, across the calendar month), maximal exercise was associated with earlier sleep onset compared to lower levels of strain and non-exercise days.

R2.2. Did you include the week following daylight saving time clock changes in your analysis? A recent study published in the Journal of Sleep Research suggests that catch-up sleep occurs after both transitions, with sleep duration slightly higher from Monday to Friday compared to before, although this effect was less pronounced in autumn (https://onlinelibrary.wiley.com/doi/10.1111/jsr.14335).

Thank you for this recommendation. To include a one-year study interval, we did include the week following both clock changes in our primary analysis. We conducted a sensitivity analysis excluding the week following the start and the week following the end of daylight saving time, and the results did not materially change (Figure 3 below). We also compared the average sleep duration during the week (Mon – Fri) before each clock change to the week (Mon – Fri) immediately after each clock change. There were no significant differences during the Autumn (before ST transition = 6.978 hours; after ST transition = 6.967 hours) or Spring (before DST transition = 6.902 hours; after DST transition = 6.905 hours) transitions in our data. As a result, we included all nights in the final analysis.

We now include this information and citation in the manuscript, in Methods, page 17:

Finally, we conducted a sensitivity analysis excluding the week following the start and the week following the end of daylight saving time, as prior research suggests catch-up sleep occurs after clock transitions.[43] The results did not change.

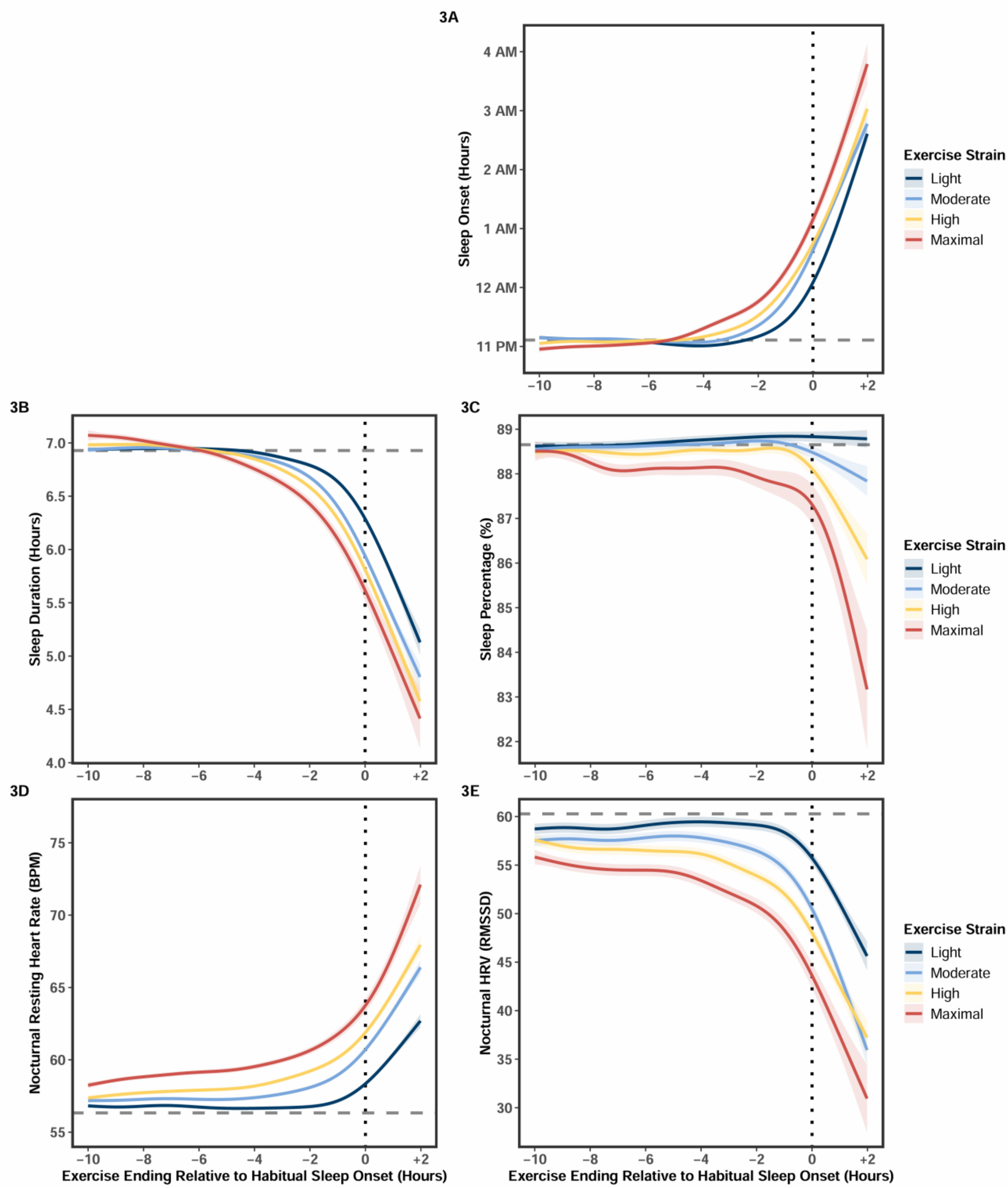

Figure 3. GAMMs demonstrating the relationship between exercise timing and **A.** sleep onset, **B.** sleep duration (in hours), **C.** sleep quality, **D.** nocturnal resting heart rate, and **E.** nocturnal HRV at different levels of exercise strain after **excluding the week following the start and the week following the end of daylight saving time**. Horizontal grey dotted line indicates the mean outcome variable on non-exercise days. Vertical black dotted line indicates habitual sleep onset.

R2.3. Are the observed associations between evening exercise timing and sleep outcomes sex-specific? Please include relevant figures in the supplement.

Thank you for this suggestion. We now include exploratory analyses examining whether our primary findings vary across gender (we collected self-reported gender identity and not biological sex). **We have included the relevant figures in the supplement (Figures S4 and S5)**, but also below for your convenience. The associations are extremely similar for men and women. We observed clear dose-dependent relationships between exercise timing and exercise strain on sleep and nocturnal autonomic activity. There were two subtle gender differences, but these differences related to averages and variance and not relationships between variables. First, on average, women went to sleep slightly earlier, slept for longer, had higher sleep percentage, and had higher nocturnal resting heart rate, consistent with the literature (Burgard et al., 2013; Reyner & Horne, 1995; Umetani et al., 1998). Second, due to having 3,844 women vs 10,845 men, the confidence intervals surrounding the smooth estimates for women were larger compared to men.

We now include this information in the Methods, page 17:

Exploratory analyses were also conducted to examine whether the findings vary across demographic groups. The primary analyses were repeated stratified across gender (Figures S4 and S5), age (Figures S6 and S7), and BMI status (Figures S8 and S9).

And Results, page 9:

The associations of interest between exercise timing and strain, and sleep and nocturnal autonomic activity did not differ across gender (Figures S4 and S5), age (Figures S6 and S7), or BMI status (Figures S8 and S9).

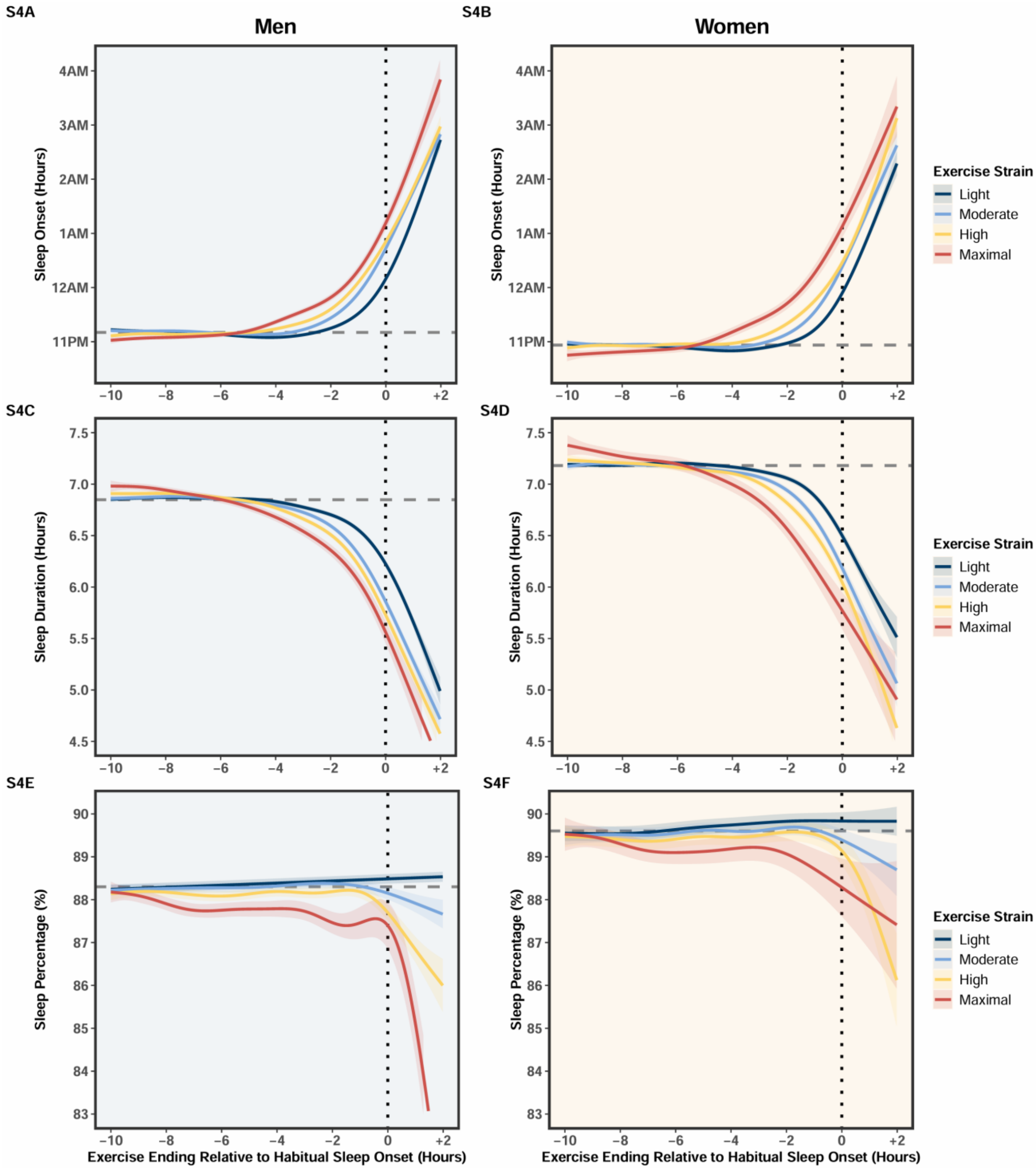

Figure S4. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** sleep onset, **C-D.** sleep duration (in hours), and **E-F.** sleep quality at different levels of exercise strain **stratified by gender** (men shaded in blue on the left; women shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

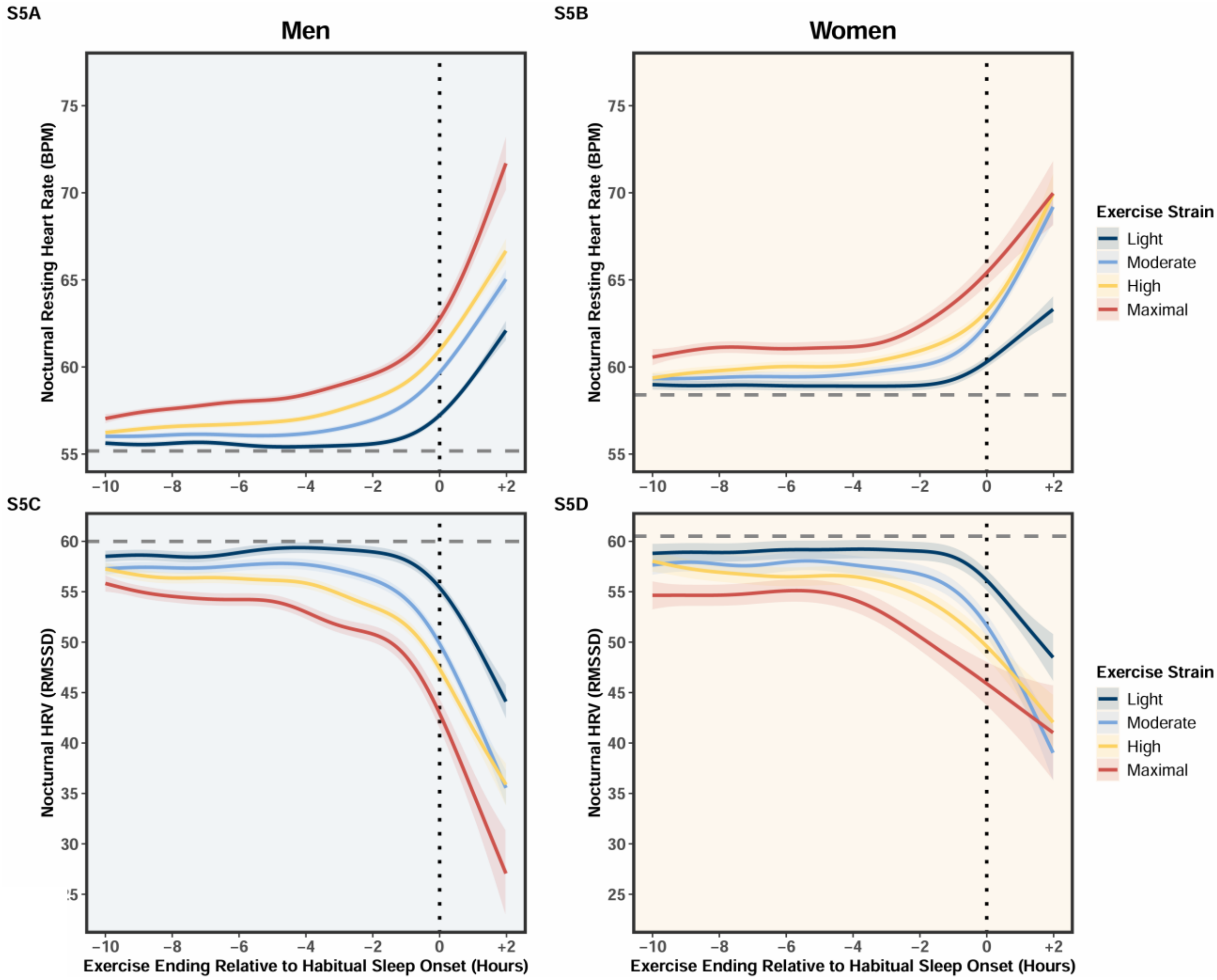

Figure S5. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** nocturnal resting heart rate, and **C-D.** nocturnal HRV at different levels of exercise strain **stratified by gender** (men shaded in blue on the left; women shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

R2.4. Are the observed associations between evening exercise timing and sleep outcomes age-specific? Consider merging the middle and upper age range groups, as the latter is too small. Again, please include relevant figures in the supplement.

We now also include exploratory analyses examining whether our primary findings vary across age group (18 to < 44 years; 44 to 87 years). **We have included the relevant figures in the supplement (Figures S6 and S7)**, but also below for your convenience. Like gender, the associations of interest were similar across age groups, with a few interesting differences. First, and consistent with the literature (Evans et al., 2021), on average, the 44–87 age group (*n* = 3,852; 26.2%) went to sleep slightly earlier, slept for slightly shorter periods, and demonstrated significantly worse sleep percentage compared to the 18–44 age group (*n* = 10,837; 73.8%). Second, and perhaps the most striking difference, the 44–87 age group had significantly lower nocturnal HRV compared to the 18–44 age group. It is well-established that HRV declines as part of normal aging (Umetani et al., 1998) and matches the age-associated decline in parasympathetic modulation of heart rate (Korkushko et al., 1991). Although the absolute magnitude of the associations between exercise timing and strain on nocturnal HRV are smaller in the 44–87 age group, the relative magnitude of these associations are similar across both groups.

We now reference and report this analysis in the manuscript and supplement.

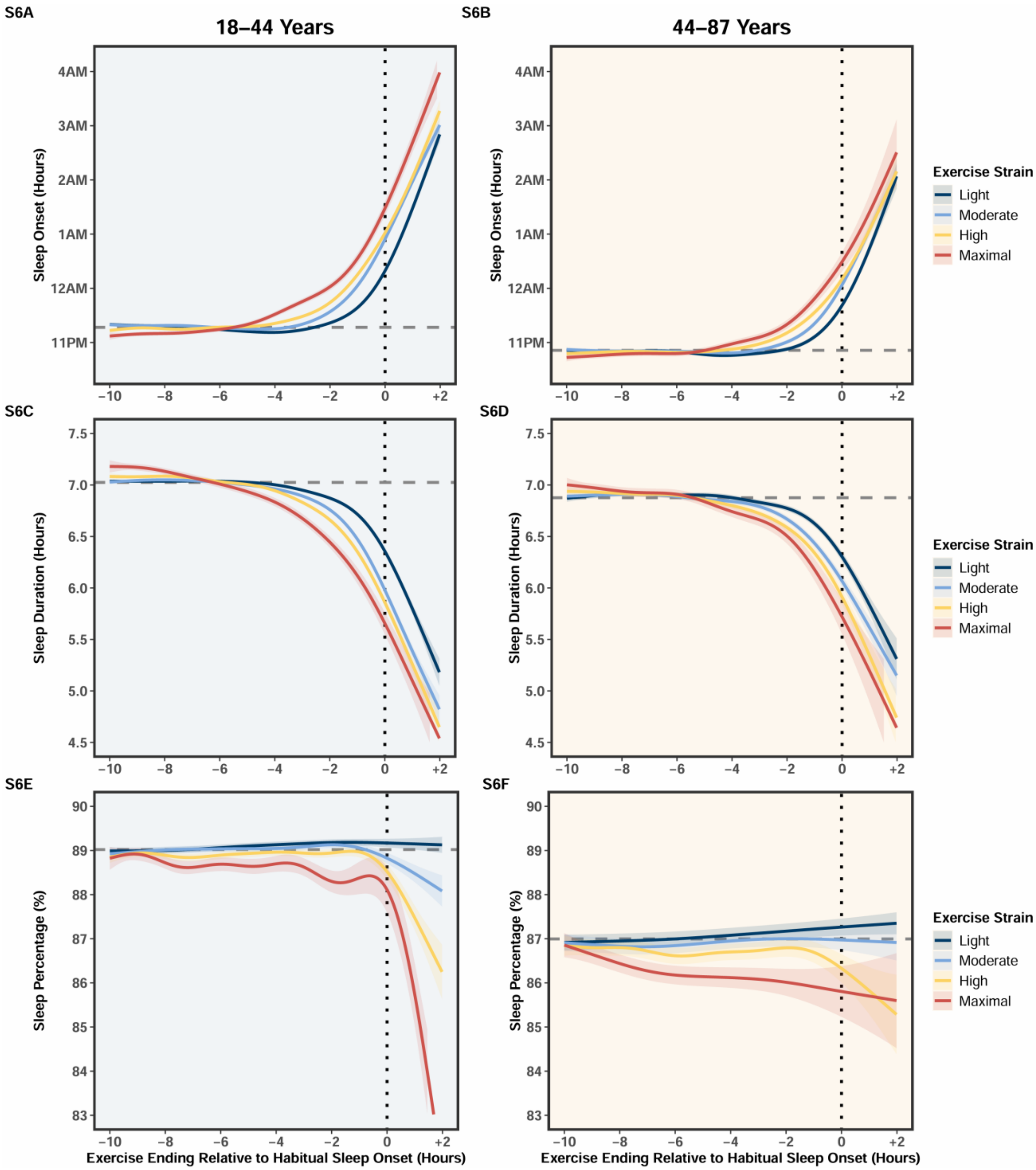

Figure S6. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** sleep onset, **C-D.** sleep duration (in hours), and **E-F.** sleep quality at different levels of exercise strain **stratified by age** (18–44 years of age shaded in blue on the left; 44–87 years of age shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

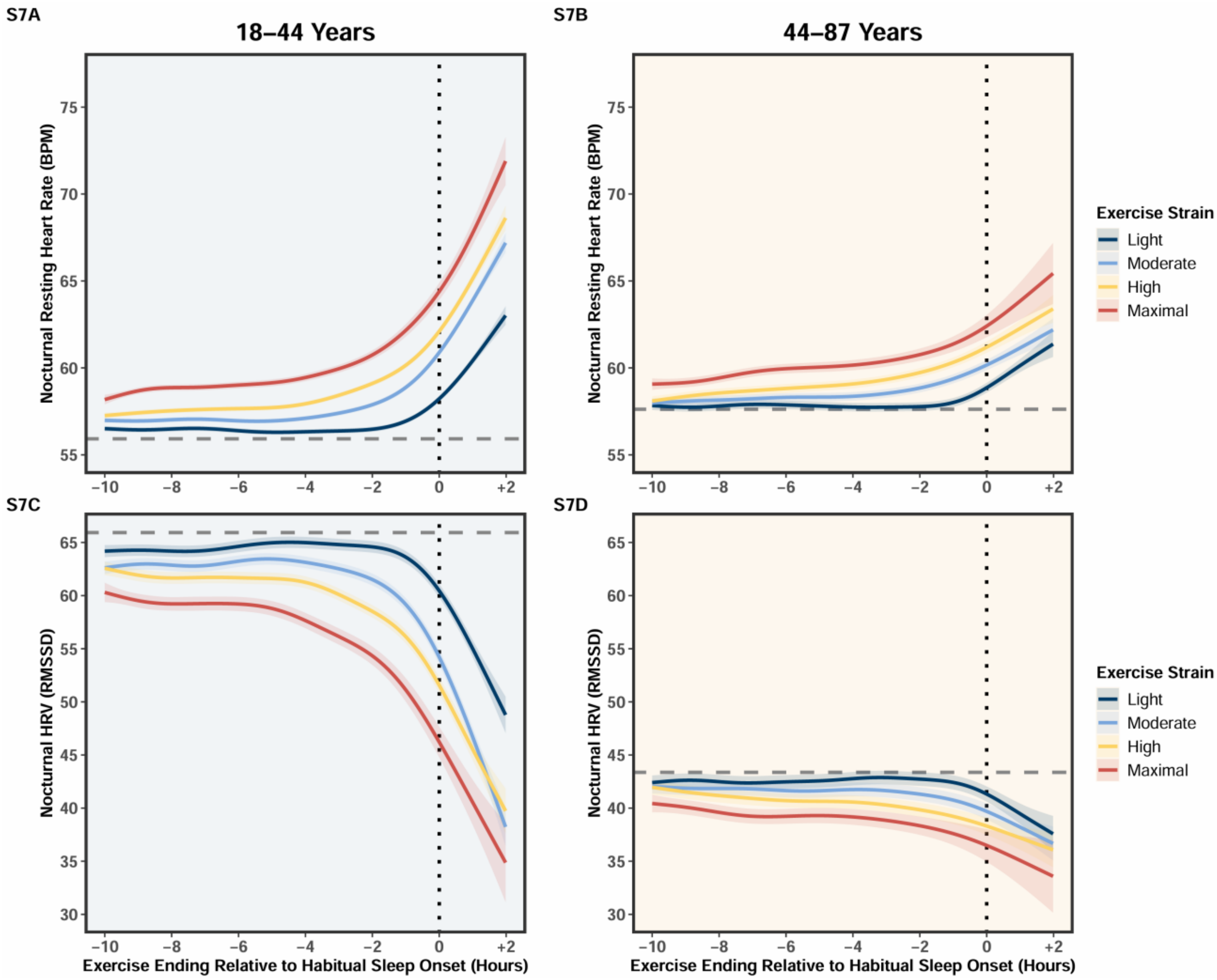

Figure S7. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** nocturnal resting heart rate, and **C-D.** nocturnal HRV at different levels of exercise strain **stratified by age** (18–44 years of age shaded in blue on the left; 44–87 years of age shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

R2.5. Given the seasonal effects on sleep, I would like to see whether the observed associations between evening exercise timing and sleep outcomes vary by season. For example, evening exercise may be more likely to occur outdoors in summer, when temperatures are higher and evenings are brighter.

This is an interesting question. Due to confidentiality considerations, we do not have access to location data and given our multinational sample we would not feel confident including seasonal analyses (as well as temperature and light exposure) in the manuscript given we would be making large assumptions about when seasons occurred for each participant.

Importantly, we did consider the influence of seasonal effects on sleep and exercise in the current analysis by quantifying habitual sleep onset timing as the participant's average sleep onset time of that calendar month (see R2.1. above for more information). The reviewer is likely correct that evening exercise may be more likely to occur outdoors in the summer when temperatures are higher and evenings are brighter. If we assume that most participants resided in the United States for the entire study period, it appears that exercises ended later and were longer in duration in the summer compared to the winter (Figure 8 below). However, when we reran the primary analyses stratified by season (American winter vs. American summer), the findings were extremely similar (Figures 9 and 10). Taken together, people may respond similarly to strenuous evening exercise in the winter and summer, even though evening exercise may occur more often, or at least later in the day, in the summer.

We now acknowledge that we did not have access to location data in the Methods, page 13:

Due to confidentiality considerations, we did not have access to location data.

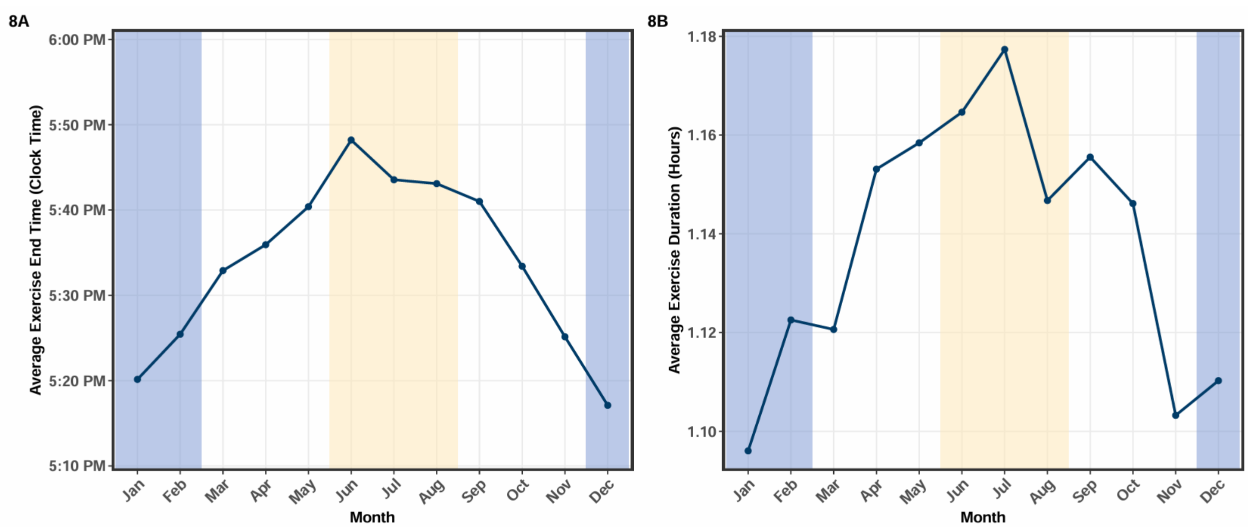

Figure 8. Monthly and seasonal breakdown of **A.** average exercise end timing, and **B.** average exercise duration. Winter months are shaded in blue. Summer months are shaded in yellow. Note seasons are based on the United States calendar.

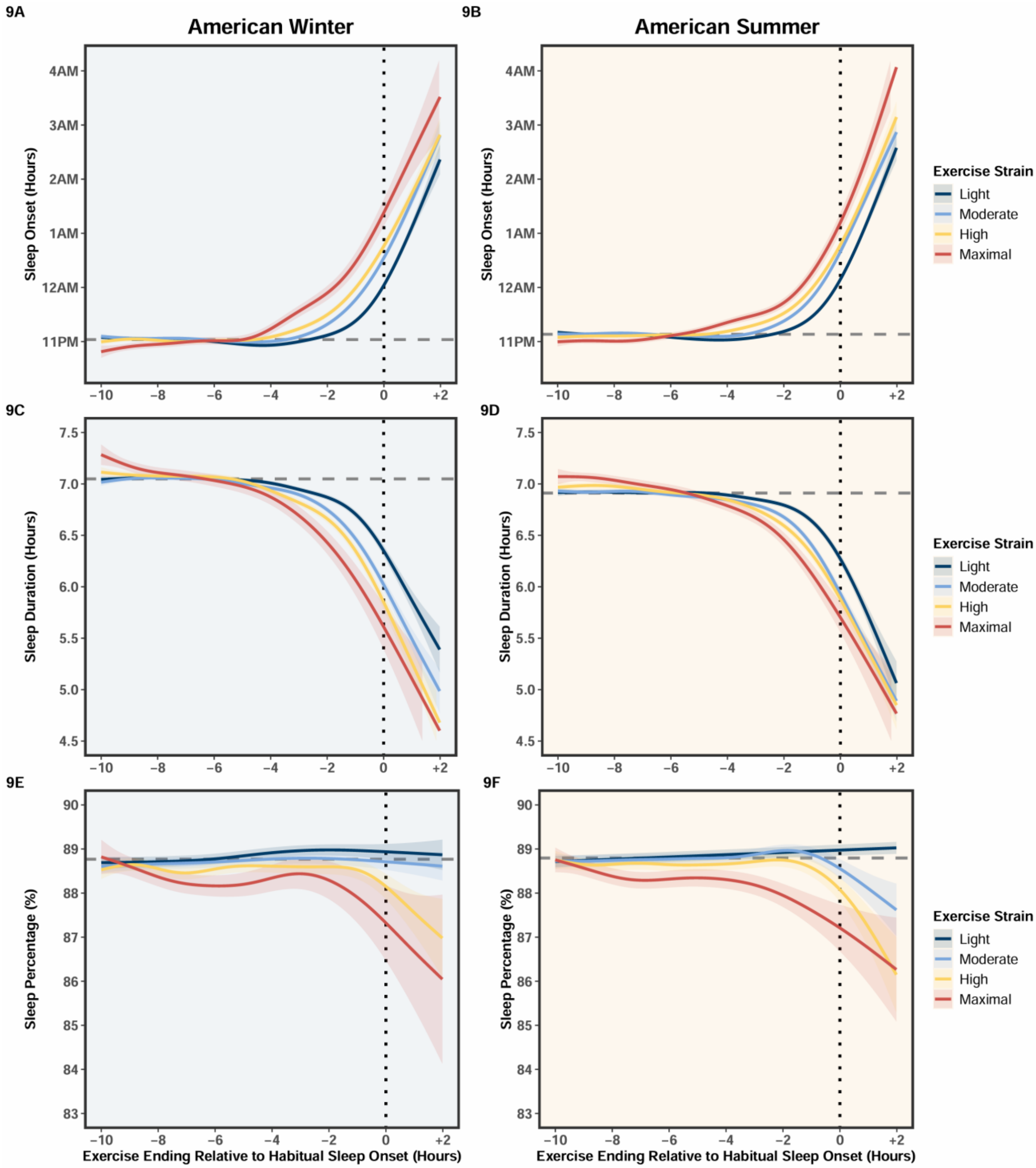

Figure 9. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** sleep onset, **C-D.** sleep duration (in hours), and **E-F.** sleep quality at different levels of exercise strain **stratified by season** (winter shaded in blue on the left; summer shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

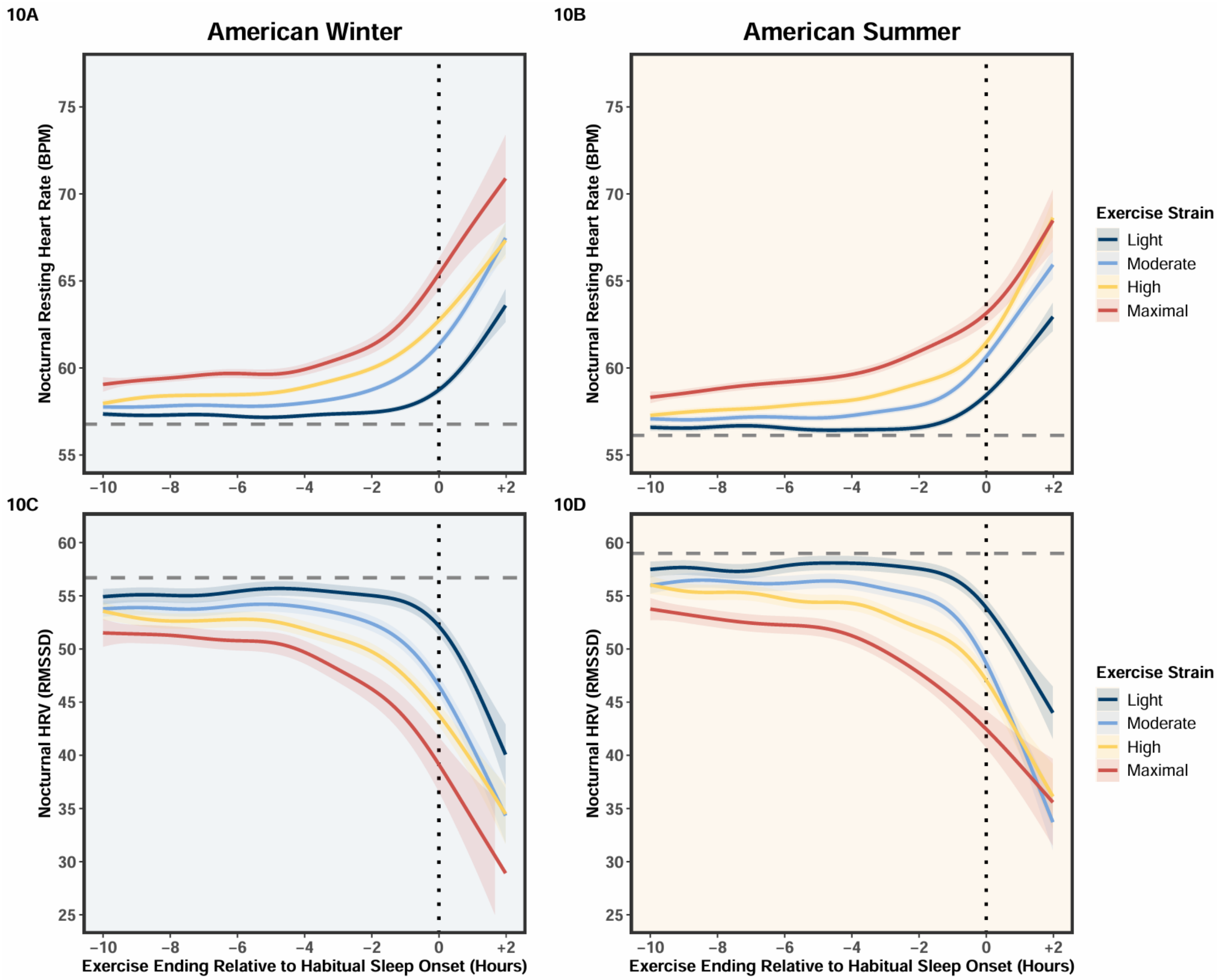

Figure 10. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** nocturnal resting heart rate, and **C-D.** nocturnal HRV at different levels of exercise strain **stratified by season** (winter shaded in blue on the left; summer shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

R2.6. Morning and early afternoon exercise, even based on smaller samples, has been proposed to advance the onset of sleep. Therefore, I wonder whether morning exercise (for which you have data, correct?) could at least partially offset some of the observed detrimental effects of evening exercise on sleep outcomes.

We investigated the question of whether morning and early afternoon exercise advanced subsequent sleep onset timing in an exploratory analysis by **including exercises concluding 10–16 hours before habitual sleep onset** (e.g., concluding between 7:00 AM and 12:00 PM for individuals with an 11:00 PM average sleep onset; Figure 10 below). We did not observe any clear or substantial sleep timing advance across exercise strain groups. Although it is certainly an interesting research question, we believe it is outside the scope of the current study and better suited for highly controlled laboratory studies. For example, future research examining the circadian phase-shifting effects of exercise could test whether the phase response varies depending on the level of exercise strain. Although research in animal models indicates higher exercise strain may induce greater phase shifts (Bobrzynska & Mrosovsky, 1998) this has yet to be tested in humans.

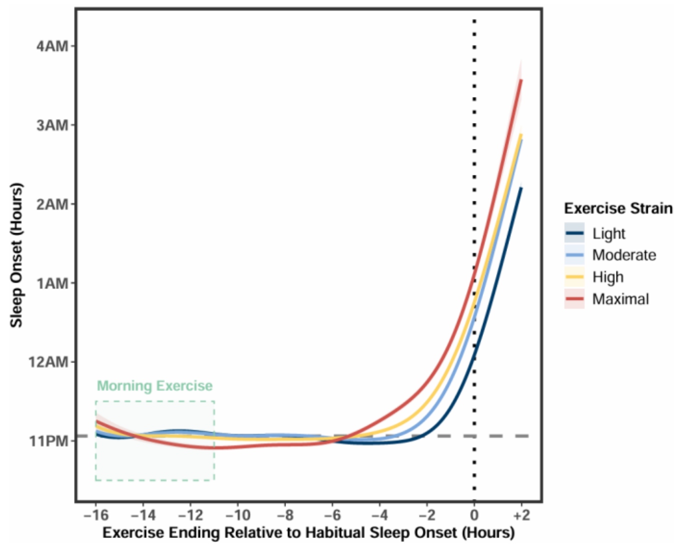

Figure 11. GAMM demonstrating the relationship between exercise timing and sleep onset timing at different levels of exercise strain. **Morning exercises (i.e., 10–16 hours before habitual sleep onset) highlighted in teal**. Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dotted line indicates the mean outcome variable on non-exercise days. Vertical black dotted line indicates habitual sleep onset.

R2.7. Please add a table to the supplement comparing basic characteristics between those included in the study and those excluded.

Thank you, we now include this table in the supplement (Table S10), and below for your convenience.

Table S10. Demographic Characteristics of Individuals Included In and Excluded From The Final Sample

Group	**Included**		**Excluded**	
Variable	**Mean ± SD**	**No. (%)**	**Mean ± SD**	**No. (%)**
Total		14,689 (73.4)		5,311 (26.6)
Gender				
Women		3,844 (26.17)		1,361 (25.6)
Men		10,845 (73.83)		3,926 (73.9)
NA				24 (0.5)
Age, y	37.89 ± 10.81		39.94 ± 10.41	
Weight, KG	80.39 ± 15.10		84.34 ± 18.80	
Height, CM	176.39 ± 9.41		176.39 ± 9.76	
Cardiovascular indices				
Resting heart rate	56.41 ± 7.16		60.52 ± 7.97	
HRV	59.95 ± 27.31		50.49 ± 23.62	

R2.8. Exercise, particularly high-intensity or maximal activities, may be more physiologically exhausting for individuals with a high BMI (e.g., ≥30 kg/m²) compared to those with normal weight. Thus, I would like to see whether the observed associations between evening exercise and sleep outcomes vary by BMI status.

We now include exploratory analyses examining whether our primary findings vary across BMI status, using the suggested criteria of ≥ 30 kg/m². **We have included the relevant figures in the supplement (Figures S8 and S9)**, but also below for your convenience. Most of our sample had a BMI below 30 (*n* = 13,127; 89.4%), which was not surprising given they are physically active and subscribe to a sleep and fitness wearable device. Only 1,562 participants (10.6%) had a BMI ≥ 30. Caution should be taken when interpreting BMI scores given that BMI does not account for muscle mass and therefore it is reasonable to expect the current sample contains athletic individuals with BMI scores above 30 but without excess body fat. Indeed, this month the Lancet Diabetes and Endocrinology Commission — which is endorsed by 75 medical organizations around the world — set out a new way to diagnose obesity (Rubino et al., 2025). This reclassifies ~20% of people with BMI above 30 who would have otherwise been diagnosed with obesity, and we imagine that number could be even higher in this sample. In any event, the associations of interest were again consistent across BMI status, with the exception of sleep percentage, which did not change as a function of exercise timing among individuals with a BMI over 30.

We agree with the reviewer that some activities may be more physiologically exhausting for individuals with a high BMI given the total mass they are moving is greater. However, the current analyses were designed to handle this assumption by quantifying exercise strain using SHRZS, which quantifies exercise intensity in a manner that is relative to each person. For example, an individual with a high BMI may find a tennis match more physiologically exhausting than an individual with a low BMI, but this increased exhaustion can—in part—be captured by a larger relative elevation in heart rate and, consequently, a higher exercise strain score. This is a key strength of the current study and might also explain why we do not find significant differences in the associations of interest across BMI status.

We now reference and report this analysis in the manuscript and supplement.

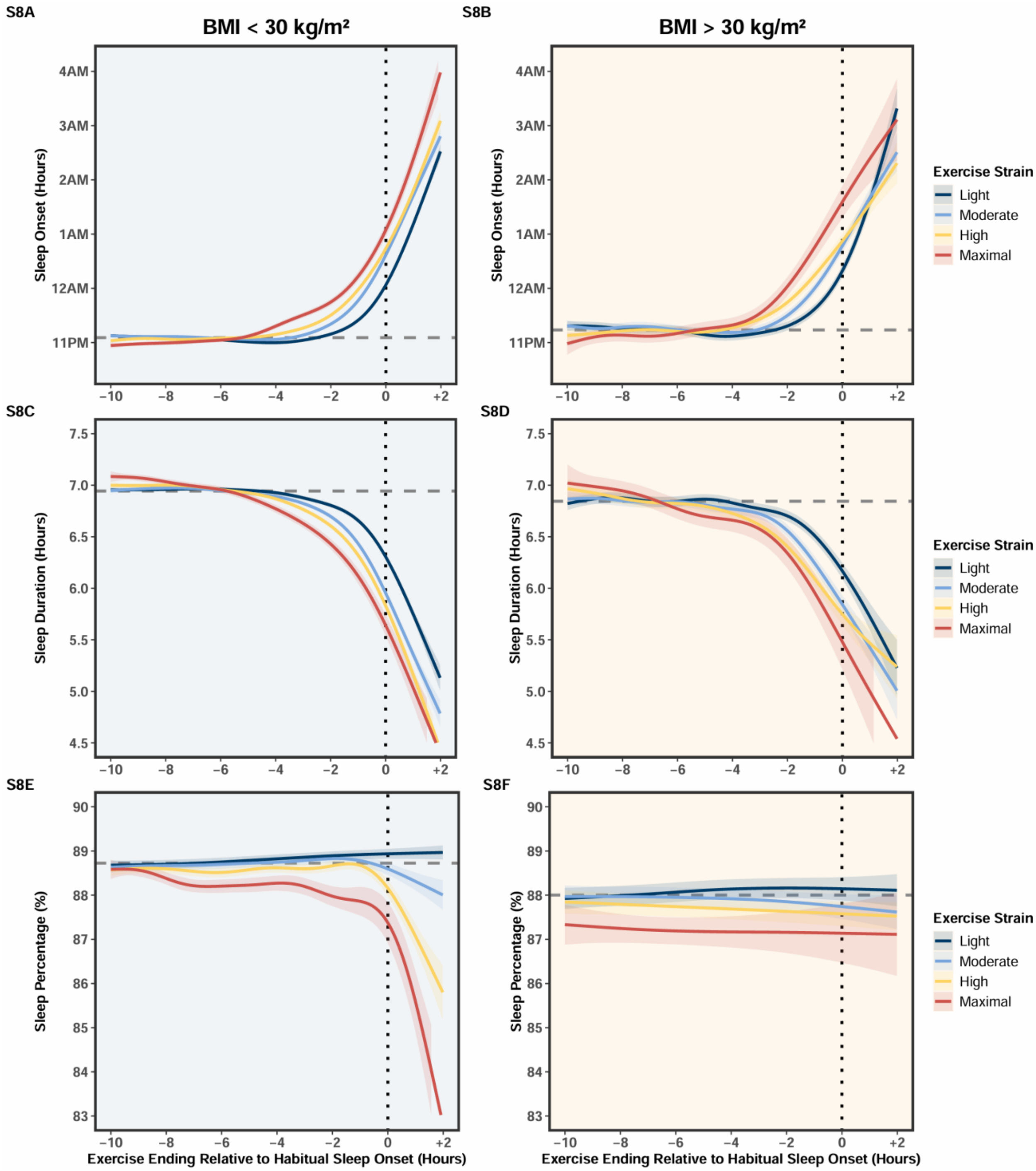

Figure S8. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** sleep onset, **C-D.** sleep duration (in hours), and **E-F.** sleep quality at different levels of exercise strain **stratified by BMI status** (BMI < 30 [kg/m²] shaded in blue on the left; BMI ≥ 30 [kg/m²] shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

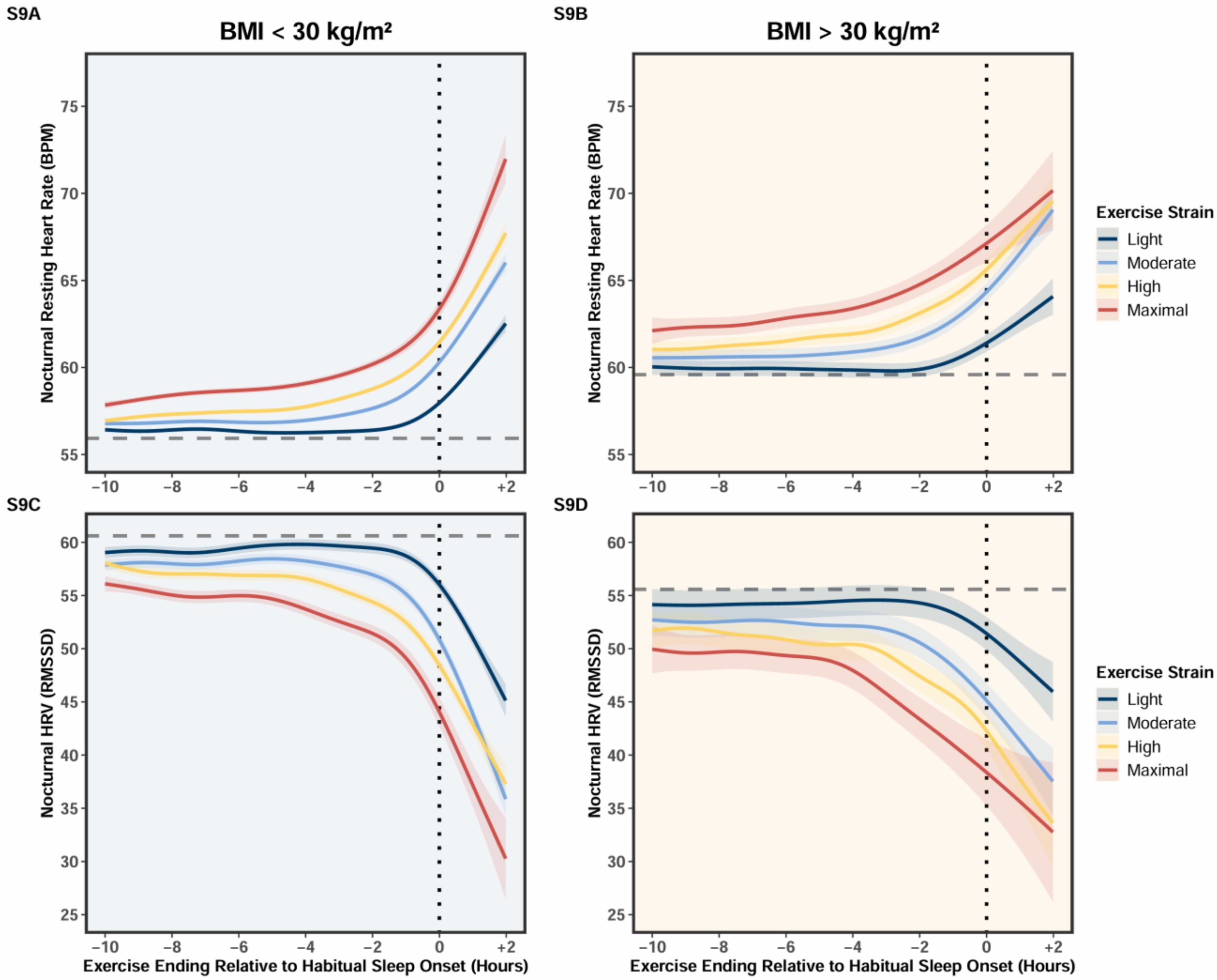

Figure S9. GAMMs demonstrating the relationship between exercise timing relative to habitual sleep onset and **A-B.** nocturnal resting heart rate, and **C-D.** nocturnal HRV at different levels of exercise strain **stratified by BMI status** (BMI < 30 [kg/m²] shaded in blue on the left; BMI ≥ 30 [kg/m²] shaded in cream on the right). Shaded area surrounding each exercise strain group line represents 95% confidence intervals. Horizontal grey dashed line indicates the mean outcome variable on no exercise days. Vertical black dotted line indicates habitual sleep onset.

References:

Atoui, S., Chevance, G., Romain, A. J., Kingsbury, C., Lachance, J. P., & Bernard, P. (2021). Daily associations between sleep and physical activity: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *57*, 101426.

Bobrzynska, K. J., & Mrosovsky, N. (1998). Phase shifting by novelty-induced running: activity dose-response curves at different circadian times. *Journal of Comparative Physiology A*, *182*, 251-258.

Borresen, J., & Lambert, M. I. (2008). Quantifying training load: a comparison of subjective and objective methods. *International Journal of Sports Physiology and Performance*, *3*(1), 16-30.

Borresen, J., & Lambert, M. I. (2009). The quantification of training load, the training response and the effect on performance. *Sports Medicine*, *39*, 779-795.

Buchheit, M., Laursen, P. B., Al Haddad, H., & Ahmaidi, S. (2009). Exercise-induced plasma volume expansion and post-exercise parasympathetic reactivation. *European Journal of Applied Physiology*, *105*, 471-481.

Burgard, S. A., & Ailshire, J. A. (2013). Gender and time for sleep among US adults. *American Sociological Review*, *78*(1), 51-69.

Edwards S. (1993). *The Heart Rate Monitor Book*. Sacramento, CA: Fleet Feet Press.

Evans, M. A., Buysse, D. J., Marsland, A. L., Wright, A. G., Foust, J., Carroll, L. W., ... & Hall, M. H. (2021). Meta-analysis of age and actigraphy-assessed sleep characteristics across the lifespan. *Sleep*, *44*(9), zsab088.

Frimpong, E., Mograss, M., Zvionow, T., & Dang-Vu, T. T. (2021). The effects of evening high-intensity exercise on sleep in healthy adults: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *60*, 101535.

Korkushko, O. V., Shatilo, V. B., Plachinda, Y. I., & ShatiloShatilo, T. V. (1991). Autonomic control of cardiac chronotropic function in man as a function of age: assessment by power spectral analysis of heart rate variability. *Journal of the Autonomic Nervous System*, *32*(3), 191-198.

Myllymäki, T., Rusko, H., Syväoja, H., Juuti, T., Kinnunen, M. L., & Kyröläinen, H. (2012). Effects of exercise intensity and duration on nocturnal heart rate variability and sleep quality. *European Journal of Applied Physiology*, *112*, 801-809.

Reyner, A., & Horne, J. A. (1995). Gender-and age-related differences in sleep determined by home-recorded sleep logs and actimetry from 400 adults. *Sleep*, *18*(2), 127-134.

Rubino, F., Cummings, D. E., Eckel, R. H., Cohen, R. V., Wilding, J. P., Brown, W. A., ... & Mingrone, G. (2025). Definition and diagnostic criteria of clinical obesity. *The Lancet Diabetes & Endocrinology*.

Scott, H., Lack, L., & Lovato, N. (2020). A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Medicine Reviews*, *49*, 101227.

Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: implications for training prescription. *Sports Medicine*, *43*, 1259-1277.

Stutz, J., Eiholzer, R., & Spengler, C. M. (2019). Effects of evening exercise on sleep in healthy participants: a systematic review and meta-analysis. *Sports Medicine*, *49*(2), 269-287.

Umetani, K., Singer, D. H., McCraty, R., & Atkinson, M. (1998). Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. *Journal of the American College of Cardiology*, *31*(3), 593-601.