scientific reports

Check for updates

OPEN

The importance of including both sexes in preclinical sleep studies and analyses

Grant S. Mannino[1,2], Tabitha R. F. Green[1], Sean M. Murphy[3], Kevin D. Donohue[4,5], Mark R. Opp[1] & Rachel K. Rowe[1]✉

A significant effort in biomedical sciences has been made to examine relationships between sex and the mechanisms underlying various disease states and behaviors, including sleep. Here, we investigated biological sex differences in sleep using male and female C57BL/6J mice (*n* = 267). Physiological parameters were recorded for 48-h using non-invasive piezoelectric cages to determine total sleep, non-rapid eye movement (NREM) sleep, rapid eye movement (REM)-like sleep, and wakefulness (WAKE). We fit hierarchical generalized linear mixed models with nonlinear time effects and found substantial sex differences in sleep. Female mice slept less overall, with less NREM sleep compared to males. Females also exhibited more REM-like sleep and WAKE and had shorter NREM sleep bout lengths. We also conducted a simulation exercise where we simulated a hypothetical treatment that altered the sleep of female mice, but not male mice. In models that included an appropriate sex by treatment interaction, a female-specific treatment response was accurately estimated when sample sizes were equal but was not detected when sample sizes were unequal, and females were underrepresented. Failure to include both sexes in experimental designs or appropriately account for sex during analysis could lead to inaccurate translational recommendations in pre-clinical sleep studies.

Keywords Female, Male, Mouse, Non-invasive, Sex differences, Sleep

All living organisms have developed physiology to promote apt circadian responses to the Earth's perpetual rotation around its axis[1]. Of these circadian processes, sleep is the most familiar, closely governing daily function and maintaining a state of homeostasis essential in preventing, managing, and ameliorating adverse health events. Thus, sleep and circadian biology have been intensively studied and applied to multiple sub-disciplines of clinical and preclinical neuroscience research. In fact, insufficient sleep has been closely linked to neurodegenerative diseases, hypertension, obesity, type-2 diabetes, cardiovascular disease, and multiple autoimmune disorders[2–7]. Such adverse effects of insufficient sleep have been largely responsible for popularizing sleep as an experimental outcome measure in preclinical research. The pace of sleep-related research continues to increase, and the number of peer-reviewed, scientific sleep journals has more than tripled since 2005[8]. Despite the recognition of sleep as a major domain of heath, inadequate sleep continues to plague human societies, warranting continued investigation into cellular and subcellular effects of altered sleep under normal physiological and pathological conditions.

Although the physiological importance of sleep is well-documented, and a body of evidence exists that indicates sex hormones affect sleep–wake behavior[9], females have been historically underrepresented in clinical sleep research[10–12]. Over the past decade, a significant effort in biomedical sciences has been made to examine relationships between sex and the mechanisms underlying various disease states and behaviors[13]. As a result, multiple reviews have highlighted the limited clinical data on differential sleep architecture between sexes[14–16]. Appropriately, greater efforts have also been made to include both male and female animals (e.g., rodents) in preclinical sleep research[13]. Unsurprisingly, animal models have proven essential to gaining further insight into the

[1]Department of Integrative Physiology, University of Colorado – Boulder, 2860 Wilderness Place, Boulder, CO 80301, USA. [2]Department of Psychology and Neuroscience, University of Colorado – Boulder, Boulder, CO, USA. [3]Cumberland Biological and Ecological Researchers, Longmont, CO, USA. [4]Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY, USA. [5]Signal Solutions, LLC, Lexington, KY, USA. ✉email: Rachel.rowe@colorado.edu

direct effects of reproductive hormones on sleep behavior, as well as the neuropathology of sleep–wake disorders. Nevertheless, when sleep is used as a physiological outcome measure in research, few studies have disaggregated sleep data based on sex or compared sleep between sexes. This failure to account for potential innate biological sex differences in sleep when analyzing data and interpreting results could cause sexually dimorphic effects to confound results and lead to flawed inferences.

Sleep–wake designation in laboratory rodents typically includes three vigilance states: wakefulness (WAKE), non-rapid eye movement (NREM) sleep, and rapid-eye-movement (REM) sleep. Although laboratory rodents are nocturnal, they are polyphasic sleepers with most sleep occurring during the light period and a smaller portion of total sleep occurring during the dark (active) period. Laboratory rodents also have much shorter sleep bouts and NREM-REM cycles than those of humans. Of the various commercially available rodent strains, the inbred C57BL/6 mouse strain is undoubtedly the most widely used in biomedical research[17]. Therefore, much of what is known about sleep outcomes following experimental manipulations has been reported from studies using this mouse strain. However, as discussed in a recent review, surprisingly few studies have been conducted that aimed to identify sex differences in the sleep of C57BL/6 mice[18].

Herein, using a large sample size of both male and female mice ($n_{Total} = 267$), we investigated biological differences in sleep between sexes to further our understanding of C57BL/6 mouse sleep architecture. We applied hierarchical statistical models with nonlinear time effects to 48-h of continuously recorded data collected under standard physiological conditions using a non-invasive piezoelectric system. Our results solidify that significant biological sex differences exist in the sleep of C57BL/6 mice and reinforce the efforts of the National Institutes of Health (NIH) to include sex as a biological variable in preclinical research.

Materials and methods

Rigor

All animal studies were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Colorado Boulder (protocol 2819) and conducted in accordance to the National Institutes of Health (NIH) guidelines for the care and use of laboratory animals. Studies are reported following the Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines[19]. In consideration of sex as a relevant biological variable, both male and female mice were used. Determination of sleep–wake behavior based on physiological parameters was done by investigators blind to the sex of the mice.

Animals

Adult (10–20 weeks-old) male and female C57BL/6J mice from 17 total sex-specific cohorts were used (males $n = 140$; females $n = 127$). This sample size ensured ≥ 0.80 power to detect a 5% difference between sexes in percentage NREM and REM sleep and percentage WAKE (Supplementary Information S1, Fig. S1.1) All mice were bred in-house from breeder pairs obtained from Jackson Laboratories (Bar Harbor, ME). Mice were group-housed until the initiation of the study protocol. Female mice were naturally cycling, and estrous stage was not controlled for as a variable. For all studies, mice were singly housed and maintained on a 12-h light:dark cycle (200 lx, cool white, fluorescent light) at an ambient temperature of 24 ± 2 °C. All mice were acclimated to non-invasive piezoelectric sleep cages for a minimum of 5 days prior to initiation of data collection. Mice were fed a normal diet of standard rodent chow, and food and water were available *ab libitum*. Following the collection of non-invasive sleep data, all mice in this study were transitioned to other studies in the laboratory, where they were manipulated, and additional data were collected. This allowed for all mice to be used in a minimum of two studies, thereby reducing the number of animals used in research.

Experimental protocol and recording parameters

To determine total sleep, physiological parameters were collected using a non-invasive piezoelectric cage system (Signal Solutions, Lexington, KY, USA), which classifies sleep–wake behavior, as previously described[20–22]. Briefly, each cage has an open bottom that allows the mouse to be placed directly on a Polyvinylidine Difluoride sensor on the cage floor. These sensors are coupled to an input differential amplifier to record pressure signals resulting from interactions with the cage floor. Pressure variations from regular breathing movements during stillness characterize sleep [about a 3 Hz, regular amplitude signal[23]], whereas signals from awake mice are of higher amplitude and display irregular spiking associated with volitional movements. In this study, the piezoelectric signals were analyzed over ten-second windows (epochs) at a two-second interval. Data collected from the sleep cages were binned at each hour using a rolling average of the percentage of recording time spent in each vigilance state[24]. Data were also binned by length of individual bout to calculate the hourly mean bout length (duration in seconds). To be considered a bout, a minimum of two consecutive epochs had to be scored as sleep (or wake). Total minutes slept within each experimental period (e.g., day 1) was also calculated.

In addition to sleep and wake behaviors, this study applied a classifier to detected sleep to discriminate between REM sleep and NREM sleep. A random forest decision tree classifier was developed to classify NREM and REM sleep based on piezo signal features related to irregularities in breathing patterns that are characteristic of changes in autonomic control during REM sleep. Specifically, cage-floor pressure signal from the animal's thorax motion- due to respiration- becomes more irregular in REM-like sleep and exhibits speed changes and low-level amplitude variations over short time intervals, which are not typically observed during NREM sleep. The classification algorithm is a part of Sleep Stats Data Explorer Version 4 and was applied in this study. The non-invasive piezoelectric data acquisition system, PiezoSleep with Sleep Stats software programs have been used to accurately determine REM and NREM in both the mouse[25] and the rat[26], and further validation in the mouse is presented in Supplementary Information S2 (Figs. S2.1 and S2.2, Tables S2.1 and S2.2). This 3-state algorithm using the piezo system recordings to determine WAKE, NREM, and REM sleep has been validated against data

collected from mice that were simultaneously recorded using EEG/EMG human-scored vigilance state classification on 4-s intervals. Automated piezoelectric scoring has been shown to distinguish WAKE, NREM, and REM sleep with over 90% specificity for all three states. Sensitivities of each state are 90%, 81%, and 66% for WAKE, NREM, and REM, respectively[25]. Notably, a prominent source of error in the piezo signal is in the discrimination of light NREM sleep– characterized by irregular breathing and moderate delta EEG power– from REM sleep due to inherent similarities in respiration. Accordingly, the majority of undetected NREM epochs are incorrectly assigned to REM, thereby causing REM sleep to be slightly overestimated.

Statistical analyses

To investigate differences between sexes within and between days for each vigilance state (total sleep, NREM, REM-like, and WAKE), we fit hierarchical generalized linear mixed models with nonlinear time effects[27,28]. We fit all statistical models in the frequentist framework using the package 'glmmTMB' in the R statistical computing environment[29,30]. Because outcome measures of interest were either percentages (e.g., percent recording time) or overdispersed counts (e.g., bout lengths and total minutes slept), we specified Beta or negative-binomial error distributions, respectively, in the models[31,32]. Total sleep time, NREM sleep time, REM-like sleep time, WAKE time, and sleep bout lengths were sequential time-series data with cyclical light–dark period trends within which observations close to each other in time are expected to be more similar than observations further apart in time [i.e., temporal dependency among observations)[33–35]]. Additionally, the mouse sleep–wake cycle is expected to have temporally varying nonlinear interactions with predictor variables and temporally varying nonlinear effects on outcome measures[35–37]. Therefore, we included three-knot and five-knot basis splines for all sleep time and sleep architecture models, respectively, to accommodate both the expected similarity among observations within a period and the nonlinear effects of time[35,38,39]. Splines are commonly used to model nonlinear effects of continuous predictors by placing knots that identify points in the data range where intervals differ, and polynomials are used to fit the data between consecutive knots[40]. To determine the appropriate number of spline knots for each sleep time and sleep bout length model, we fit models with two to seven spline knots, conducted model selection using Akaike's Information Criterion corrected for small sample size (AIC_c), and produced estimates from the top-ranked, most parsimonious model that best described the data for each outcome[41,42].

For sleep time and bout lengths, we subdivided the data by vigilance state (total sleep, REM-like sleep, NREM sleep, and WAKE) and fit sleep state-specific models with a three-way interaction among the categorical variables sex and day, and the continuous variable ZT (time) with a basis spline[35–37]. For percent of time spent in a vigilance state, we also subdivided the data by vigilance state and fit state-specific models, but we included a two-way interaction between sex and time point, disregarding ZT (i.e., total minutes slept were derived by summing across all ZT within each day). Because the data were sourced from animals in 17 different cohorts, we included random intercepts in all models for individual animals that were nested within their respective cohort; this hierarchical structure accounted for both the dependency of repeated observations for the same individual across time as well as the inherent clustering of animals within cohorts and potential variation among those cohorts[27,43]. We based statistical inferences on a combination of coefficient estimates (β), differences between estimated conditional means (Δ) and associated 95% confidence intervals, and p-values following Tukey's adjustments for multiple comparisons[44], whereas biological inferences were based on standardized effect sizes (d)[45], all of which we obtained using the R packages 'emmeans' and 'ggeffects'[35,46]. Effect sizes were classified as small ($d = 0.10–0.20$), medium ($d = 0.21–0.80$). and large ($d > 0.8$)[47].

Simulation exercise

Simulation is an invaluable tool for evaluating the effectiveness of study designs, experimental methods, and statistical analysis approaches. Simulation is also particularly useful for testing assumptions and quantifying the bias that may result from those characteristics of studies[48]. We conducted a simulation exercise for two reasons: first, to test the validity of our findings from the empirical data, and second, to evaluate the potential consequences of ignoring a true sex difference in a treatment versus control study based on a 'mixed sex' sample of mice[48]. For this simulation exercise, the treatment represents a hypothetical pharmacological compound that has a sex-specific effect on sleep. The hypothetical treatment significantly increases sleep in female mice, but not male mice.

Simulations were based on pseudo-randomly generated data for total minutes slept over a 24-h period. To generate simulated data, we used the known probability distribution of total minutes slept that was generated from our empirical data presented in this study. Specifically, we simulated pseudo-random data from the negative-binomial distribution (the observed distribution of total minutes slept), using the 'rnegbin()' function in the R package MASS[49]. For the simulations, two study designs were used. In one, male and female mice were given a hypothetical compound that had a sex-specific treatment effect in which only female sleep was altered. This simulated study design included 200 total mice with equal numbers for each sex ($n = 100$ females and 100 males). These group sizes were subdivided equally into control ($n = 50$ females and 50 males) and hypothetical treatment ($n = 50$ females and 50 males) groups. We specified the following true means ($\bar{x}$) and variances (σ^2) for the distributions of each group to simulate data from: 1) Female Control: $\bar{x} = 615$ min, $\sigma^2 = 3137$; 2) Female Treatment: $\bar{x} = 645$ min, $\sigma^2 = 3419$; 3) Male Control: $\bar{x} = 665$ min, $\sigma^2 = 3613$; and 4) Male Treatment: $\bar{x} = 670$ min, $\sigma^2 = 3663$ (Supplementary Information S3, Fig. S3.1). Those distributions reflected a true biological difference in sleep between female and male controls, such that females exhibited significantly less sleep than males (similar to our empirical data); a true difference between the female control and female treatment groups, such that the female treatment group slept significantly more than the female control group (i.e., a female-specific treatment effect following the administration of a hypothetical treatment); and no differences between the male control and male treatment groups or the female treatment and male treatment groups.

In the second simulated study design, we conducted a similar but separate exercise that reflected a pre-clinical sleep research study that had unequal sample sizes with fewer female mice than male mice (i.e., uneven sample sizes; Supplementary Information S3, Fig. S3.2). In this simulation, male and female mice were given a hypothetical compound that had a sex-specific treatment effect in which only sleep of female mice was altered. This simulated study design included 200 total mice with unequal sample sizes for each sex ($n = 50$ females and 150 males). These group sizes were subdivided equally into control ($n = 25$ females and 75 males) and hypothetical compound ($n = 25$ females and 75 males) groups.

In both study designs, we investigated if a treatment effect was found when sleep data were disaggregated by sex, and when sleep data were combined in 'mixed sex' groups. For both the equal and unequal sample size scenarios, we fit two generalized linear models with negative-binomial error distributions to the simulated data using the R package glmmTMB; one model included a sex × treatment fixed effects interaction that appropriately accounted for the observed sex difference in sleep (used to represent sex differences in sleep for the control group in the simulations), and the female-specific treatment effect on total minutes slept. The other model only included a treatment fixed effect and disregarded both the baseline sex difference and the female-specific treatment effect (i.e., a conventional 'mixed sex' analysis that combined sexes together for inference). We evaluated the results from each model by calculating point estimate relative bias and 95% confidence interval coverage of the true (simulated) means specified above, using the R package SimDesign[48,50].

Results

Sleep–wake behavior was similar between day 1 and day 2

Sleep–wake behavior for all outcome measures was similar between recording day 1 and day 2. No statistically significant differences existed between days, for either sex, for any outcome measure assessed in this study (Table 1).

Sex differences in hourly sleep–wake behavior

Hourly percent sleep was assessed, and female mice exhibited significantly less total sleep, determined with the two-state discriminator, than did males during the first five hours of the light period (ZT 0–4; $p = 0.0001–0.03$; Fig. 1A) and the last nine hours of the dark period (ZT 15–23; $p = 0.0001–0.02$). Total hourly percent sleep was similar between sexes within each day from ZT = 5 to ZT = 14 ($p = 0.05–0.93$), i.e., during the latter half of the light period and the beginning of the dark period. Standardized effect sizes for supported differences in total hourly percent sleep between sexes were small ($d \leq 0.11$).

NREM hourly percent sleep did not differ between days for either sex ($\beta_{Sex \times Day} = -0.01$, $p = 0.95$; Fig. 1C). Within each day, females exhibited significantly less hourly percent NREM sleep than males across the entire light (ZT 0–11; $p = 0.0001–0.04$; Fig. 1B) and dark (ZT 12–23; $p = 0.0001–0.02$; Fig. 1B) periods. Standardized effect sizes for supported differences in NREM hourly percent sleep between sexes were small at all ZTs ($d \leq 0.10$).

Female mice exhibited significantly more REM-like hourly percent sleep than male mice across the entire light period (ZT 0–11; $p = 0.0001–0.02$; Fig. 1C). In contrast, REM-like hourly percent sleep was similar between sexes during the entire dark period within each day (ZT 12–23; $p = 0.07–0.97$; Fig. 1C). Standardized effect sizes for supported differences in REM-like hourly percent sleep between sexes were small at all ZTs ($d \leq 0.01$).

Female mice exhibited significantly more hourly percent WAKE than male mice during the first five hours of the light period (ZT 0–4; $p = 0.0003–0.04$; Fig. 1D) and the last ten hours of the dark period (ZT 14–23; $p = 0.0001–0.04$). Hourly percent WAKE was similar between sexes within each day from ZT 5 to ZT 13 ($p = 0.05–0.99$). Standardized effect sizes for supported differences in hourly percent WAKE between sexes were small at all ZT ($d \leq 0.10$).

Outcome measure	$\beta_{Sex \times Day}$[a]	*P*-value[b]	Figure[c]
Total sleep (%)	−0.02	0.87	1A
NREM sleep (%)	−0.01	0.95	1B
REM sleep (%)	−0.09	0.18	1C
WAKE (%)	0.03	0.82	1D
Total sleep (min)	−0.01	0.22	3A
NREM sleep (min)	0.004	0.77	3B
REM sleep (min)	0.04	0.07	3C
WAKE (min)	0.0006	0.94	3D
Total sleep bout length (sec)	−0.04	0.59	5A
NREM sleep bout length (sec)	0.10	0.16	5B
REM sleep bout length (sec)	0.003	0.95	5C
WAKE bout length (sec)	0.03	0.85	5D

Table 1. Similarities in sleep–wake behavior between day 1 and day 2. No statistically significant ($p < 0.05$) differences existed between day 1 and day 2, for either sex, for any outcome measures assessed in this study. [a]Coefficient estimate for the sex × day interaction from the fitted model. [b]Probability of obtaining results at least as extreme as the results observed if the null hypothesis is correct. [c]Corresponding figure to which the coefficient estimate applies.

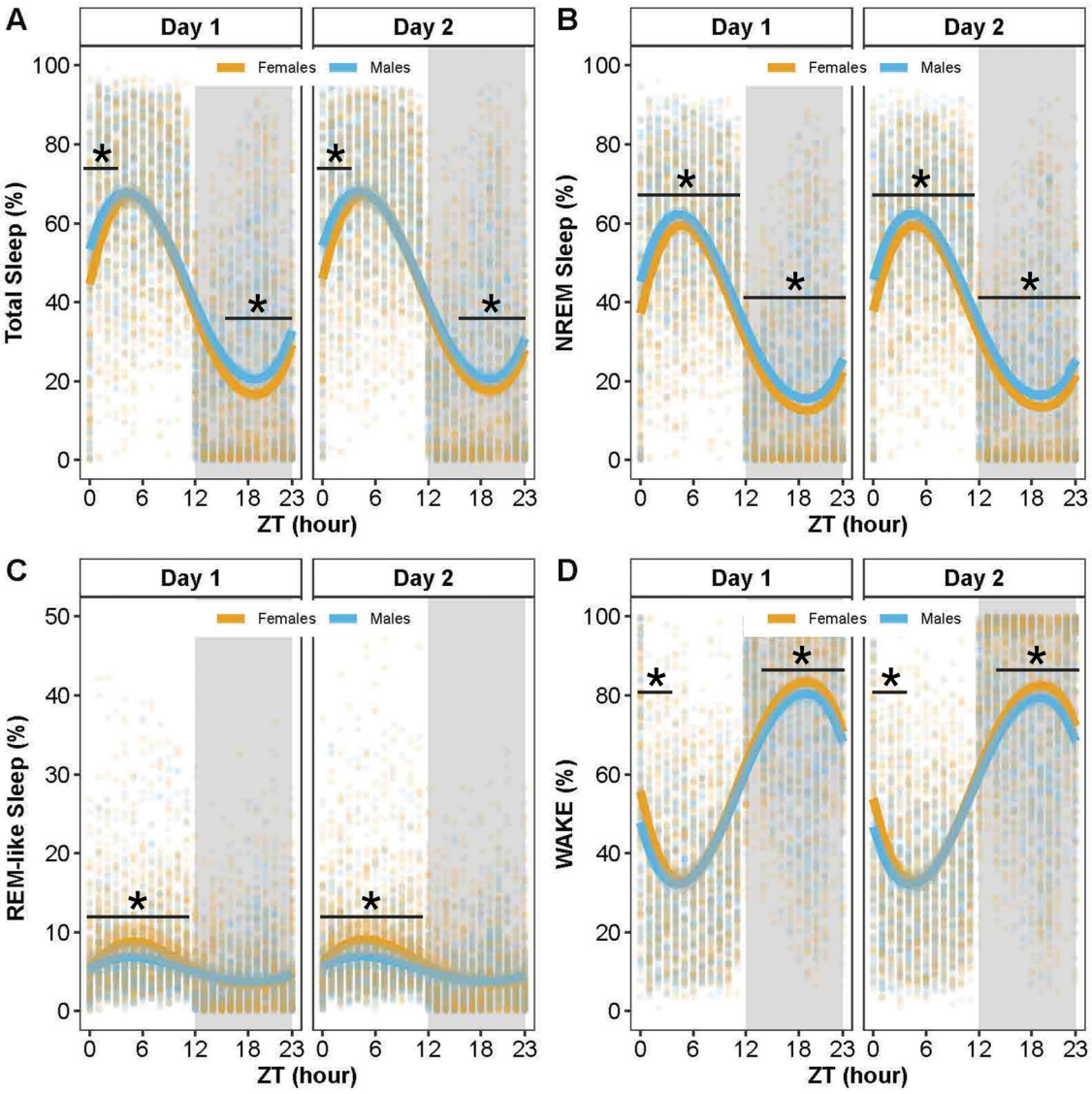

Fig. 1. Percent of time spent in vigilance states differed between sexes during both the light and dark periods across all sleep types. Statistically significant sex differences ($p < 0.05$) existed for percent of time spent in (**A**) Total sleep, (**B**) NREM sleep, (**C**) REM-like sleep, and (**D**) WAKE. Results shown are predicted conditional effects point estimates and 95% confidence intervals estimated by hierarchical models with Beta error distributions; background points denote the observed data values.

We also assessed variation in the hourly sleep percentages among cohorts used in this study (Fig. 2). Each model accounted for variation among cohorts via random intercepts. Variation in total sleep time, NREM sleep, and WAKE among cohorts was nominal and not statistically significant. Variation in REM-like sleep among cohorts was nominal and not statistically significant, except for cohort 8 that exhibited a positive effect (i.e., cohort 8 exhibited more hourly percent REM-like sleep; Fig. 2C).

Sex differences in daily sleep–wake behavior

Data were also binned over 24-h periods to investigate minutes of total sleep, determined with the two-state algorithm, and minutes spent in each vigilance state, determined from the novel three-state algorithm. Female mice exhibited less daily total sleep ($p < 0.0001$; Fig. 3A) and less daily NREM sleep time ($p < 0.0001$; Fig. 3B) than male mice, although the standardized effect sizes were small ($d \leq 0.06$). Female mice exhibited more daily REM-like sleep time than male mice ($p = 0.01$–0.04; Fig. 3C), and the standardized effect sizes were large ($d = 1.72$–2.23). Within each day, female mice exhibited significantly more minutes of WAKE than did male mice ($p < 0.0001$; Fig. 3D), although the standardized effect sizes were small ($d = 0.07$).

We also assessed variation in the daily sleep–wake behavior among cohorts used in this study (Fig. 4). Each model accounted for variation among cohorts via random intercepts. Variation in daily total sleep time, NREM sleep time, and WAKE time was nominal and not statistically significant. Similar to hourly percent sleep, the

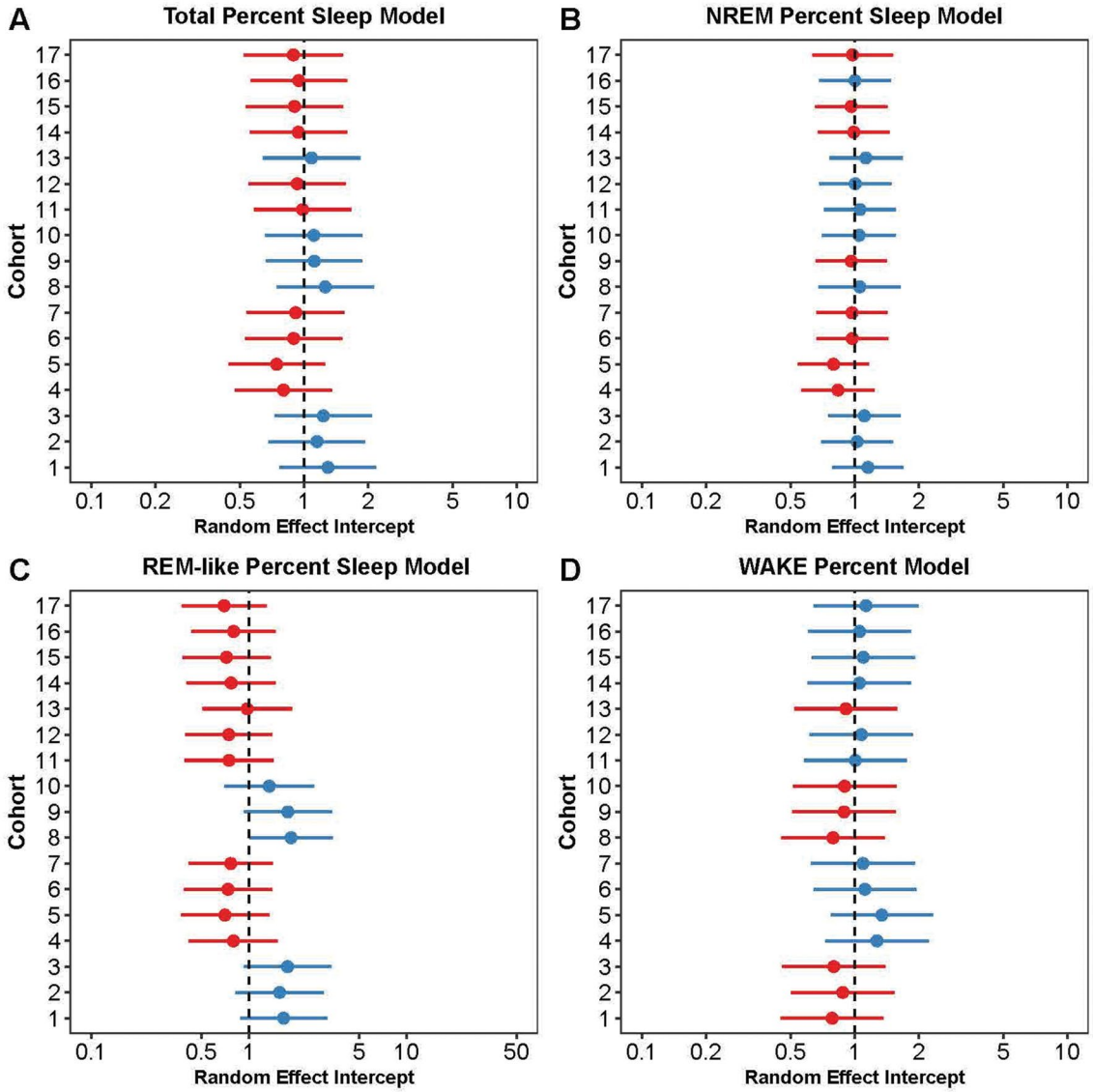

Fig. 2. Percent of time spent in vigilance states nominally varied among cohorts across all sleep types. Estimated cohort-specific random intercepts for percent of time spent in (**A**) Total sleep, (**B**) NREM sleep, (**C**) REM-like sleep, and (**D**) WAKE. Results shown are point estimates and 95% confidence intervals estimated by hierarchical models with Beta error distributions and random intercepts for individual animals nested within cohorts; blue values denote positive effects, whereas red values denote negative effects.

variation in daily REM-like sleep among cohorts, was nominal and not statistically significant, except for cohort 8 that exhibited a positive effect (i.e., cohort 8 exhibited more REM-like sleep over a 24-h period; Fig. 4C).

Sex differences in sleep architecture

We assessed the mean bout lengths of total sleep, determined with the two-state algorithm, and of each vigilance state, determined from the novel three-state algorithm. For total sleep, female mice exhibited significantly shorter bout lengths than male mice across the entire light period (ZT 0–11; $p = 0.0001–0.01$; Fig. 5A), and between ZT 17 and ZT 22 during the dark period ($p = 0.0001–0.007$). In contrast, bout lengths for total sleep were similar between sexes from ZT 12 to ZT 16 within each day ($p = 0.18–0.79$; Fig. 5A). Standardized effect sizes for supported differences in total bout lengths between sexes were small at all ZTs ($d \leq 0.01$).

Female mice had shorter NREM sleep bout lengths than male mice during the light period (ZT 0–11; $p < 0.0001$; Fig. 5B) and during most of the dark period (ZT 12–22; $p < 0.0001$), except at ZT = 23 ($p = 0.89$). Standardized effect sizes for sex differences in NREM sleep bout lengths were small at all time points ($d \leq 0.02$).

REM-like sleep bout lengths were similar between sexes across the entire light period (ZT 0–11; ZT 0–11; $p = 0.05–0.89$; Fig. 5C) and during the last seven hours of the dark period (ZT 17–23; $p = 0.07–0.87$). In contrast, female mice exhibited significantly longer REM-like sleep bouts during the first five hours of the dark period (ZT 12–17; $p = 0.0002–0.02$; Fig. 5C). Standardized effect sizes for sex differences in REM-like sleep bout lengths were small to medium ($d = 0.15–0.28$).

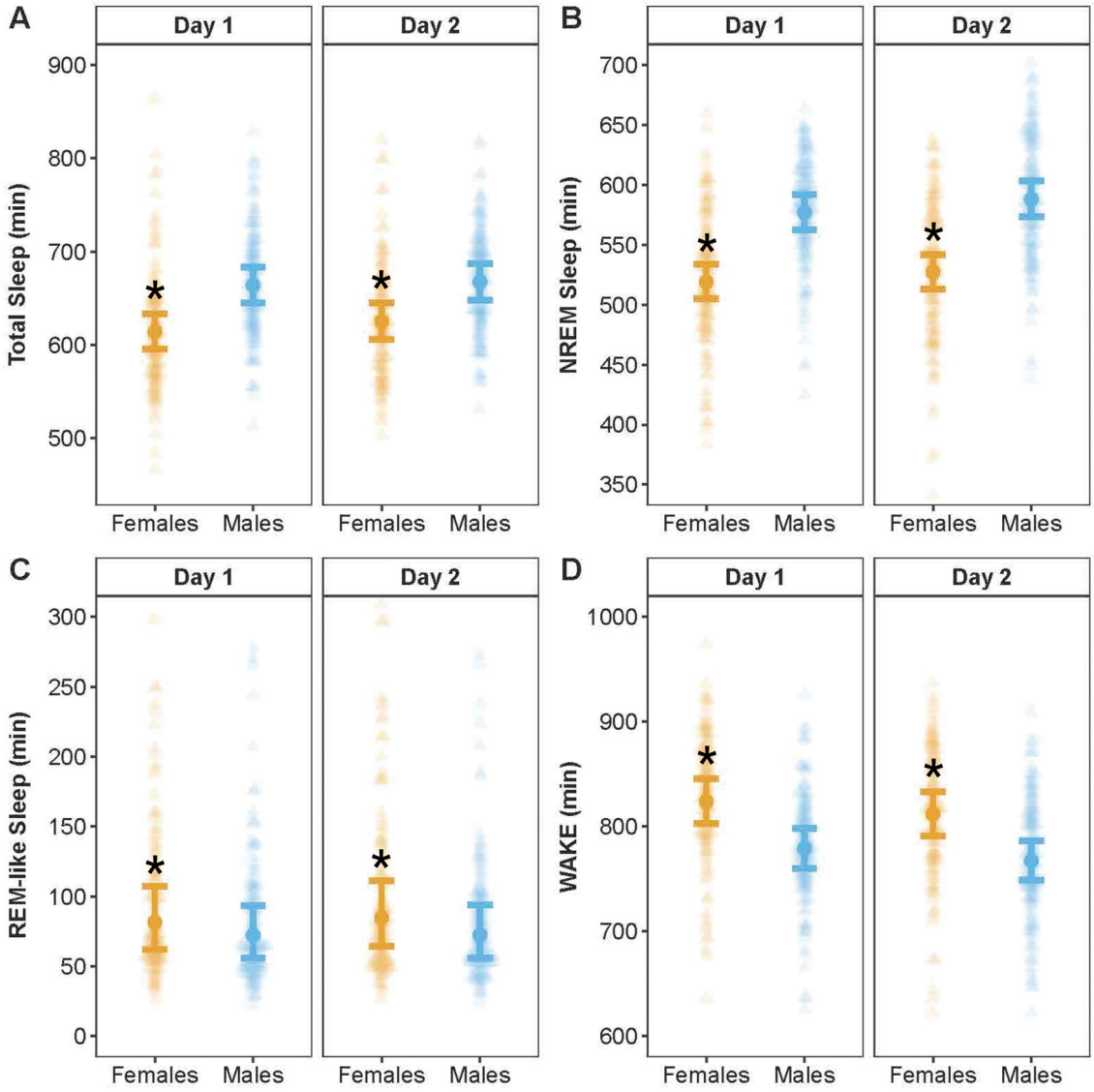

Fig. 3. Minutes of sleep and wake differed between sexes during both the light and dark periods across all vigilance states. Statistically significant sex differences ($p < 0.05$) existed for minutes spent in (**A**) total sleep, (**B**) NREM, (**C**) REM, and (**D**) WAKE. Results shown are predicted conditional effects point estimates and 95% confidence intervals estimated by hierarchical models with negative-binomial error distributions; background points denote the observed data values.

WAKE bout lengths were similar between sexes during the light period (ZT 0–11; $p = 0.05$–0.96; Fig. 5D) but differed between sexes during the dark period (ZT 12–23; $p = 0.0001$–0.03), except at ZT = 16 and ZT = 22 ($p = 0.11$–0.85). Female mice exhibited longer WAKE bouts during the dark period. Standardized effect sizes for sex differences in WAKE bout length were small at all time points analyzed ($d \leq 0.01$).

Variation in total bout lengths among cohorts, which the model accounted for via random intercepts, was nominal and not statistically significant, except for cohort 4 that exhibited a strong positive effect (Fig. 6A). Variation in NREM bout lengths (Fig. 6B), REM-like bout lengths (Fig. 6C), and WAKE bout lengths (Fig. 6D) was nominal and not statistically significant.

Simulation exercise

In the equal sample sizes scenario, the statistical model that included a sex × treatment interaction accurately estimated differences in the simulated sleep time between sexes ($p = 0.0001$) and the true female-specific treatment response ($p = 0.03$; Fig. 7A). All sex × treatment-specific 95% confidence intervals estimated by that model included the true means, and marginal mean point estimates were nominally biased relative to the true means (bias = 3.39%). In contrast, the statistical model that only included a treatment fixed effect (i.e., a 'mixed sex' analysis) failed to detect the true effects ($p = 0.23$; Fig. 7B). Only the treatment-specific 95% confidence interval estimated by that model included a true mean, the female-specific treatment mean, whereas all other true means

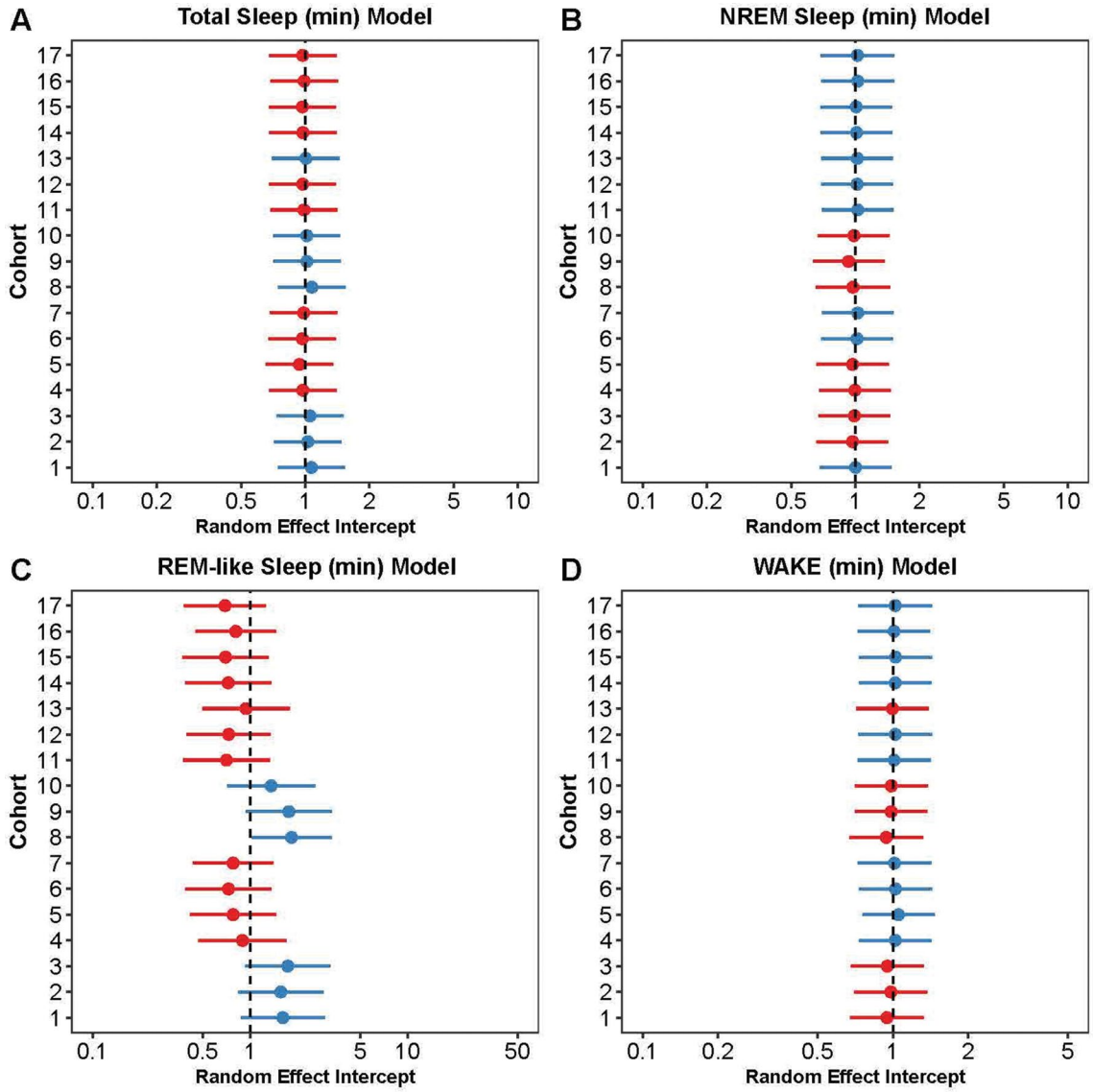

Fig. 4. REM-like sleep nominally varied among cohorts. Estimated cohort-specific random intercepts for minutes spent in (**A**) Total sleep, (**B**) NREM, (**C**) REM, and (**D**) WAKE. Results shown are point estimates and 95% confidence intervals estimated by hierarchical models with negative-binomial error distributions and random intercepts for individual animals nested within cohorts; blue values denote positive effects, whereas red values denote negative effects.

were not covered by the confidence intervals. The average bias between the marginal mean point estimates from this model and the true means was 25.00% and 12.50% for the control and treatment groups, respectively.

In the unequal sample size scenario, although the statistical model that included a sex × treatment interaction accurately estimated the true means, the greater uncertainty induced by the smaller female sample size resulted in the model not detecting the true significant baseline difference in total minutes slept or the true female-specific treatment response ($p \geq 0.27$; Fig. 8A). All sex × treatment-specific 95% confidence intervals estimated by that model included the true means, and marginal mean point estimates were nominally biased relative to the true means (bias = 2.61%); however, 95% confidence intervals for the female-specific estimates were too wide to detect a statistically significant difference. The model that only included a treatment fixed effect (i.e., a 'mixed sex' analysis) failed to detect the true effects ($p = 0.46$; Fig. 8B). Because of the greater uncertainty induced by the smaller female sample size, confidence intervals were wider, and the treatment-specific 95% confidence interval estimated by that model included two of the four true means. The average bias between the marginal mean point estimates from this model and the true means was 28.00% and 15.50% for the control and treatment groups, respectively.

Discussion

During the last 20 years, the importance of sleep to public health has become increasingly recognized[51–54]. Recent reviews have highlighted the bidirectional relationship between sleep and disease, where insufficient/disrupted sleep is both a risk factor and an outcome of various conditions[7,51,55,56]. As new pre-clinical models are developed,

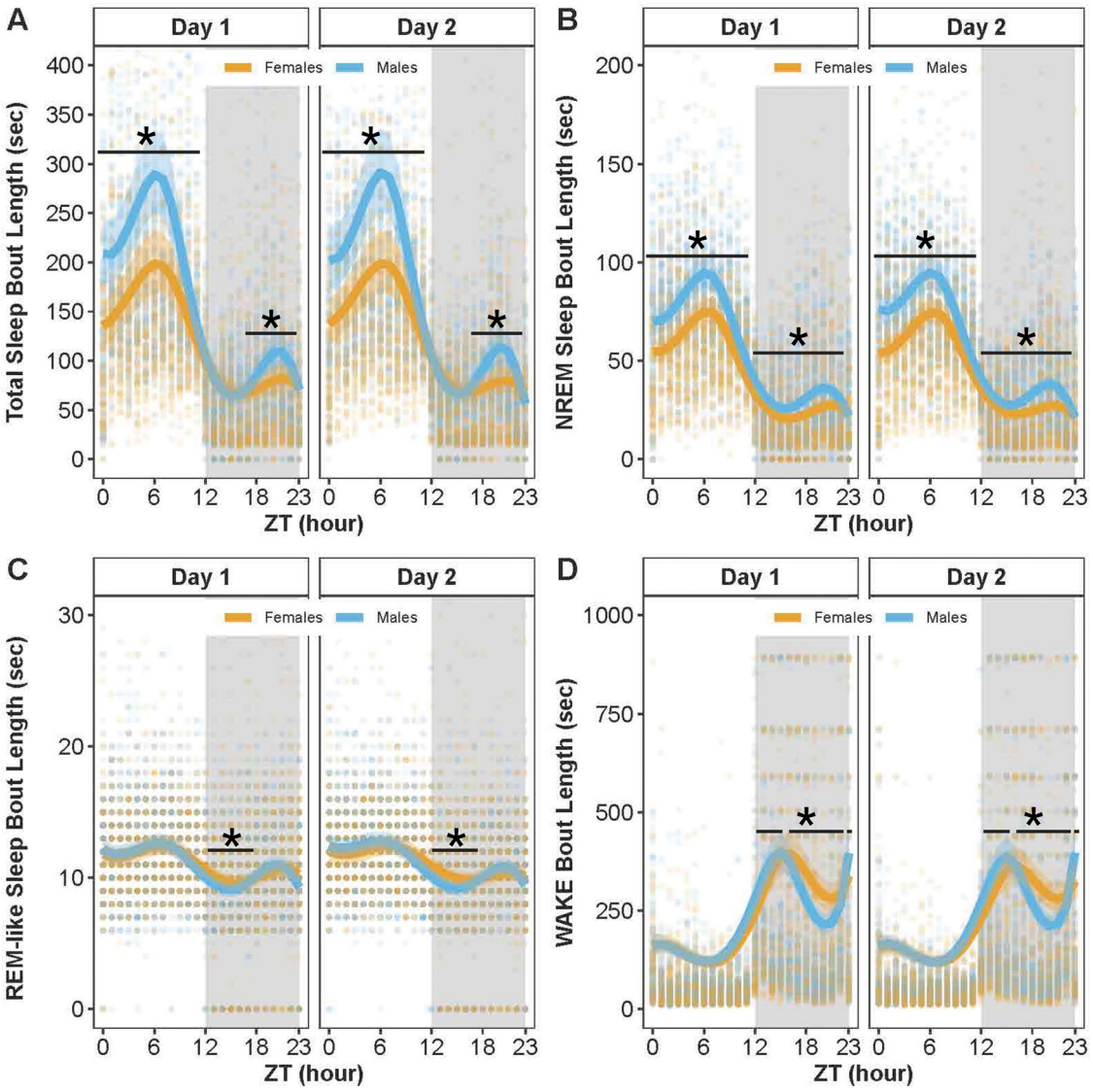

Fig. 5. Bout lengths differed between sexes during both the light and dark periods across all vigilance states. Statistically significant sex differences ($p < 0.05$) existed for (**A**) Total sleep, (**B**) NREM, (**C**) REM, and (**D**) WAKE bout lengths. Results shown are predicted conditional effects point estimates and 95% confidence intervals estimated by hierarchical models with negative-binomial error distributions; background points denote the observed data values.

determination of sleep–wake behavior and analysis of sleep architecture presents challenges for research from multi-disciplinary backgrounds.

In the current study, we found notable sex differences in the sleep–wake behavior and sleep architecture of C57BL/6J mice. Female mice slept less overall, specifically during the dark period, with less NREM sleep compared to males. Females also exhibited more total REM-like sleep and more WAKE compared to males. Sex differences have been reported in baseline sleep and the homeostatic regulation of sleep in C57BL/6J mice[57]. Similar to our current findings, other studies also demonstrated that female mice have more wakefulness and less NREM sleep, specifically during the dark period, compared to males[58]. However, previous reports on sex differences in sleep of mice generally contain data from relatively few animals without specific information with respect to whether analyses included assessment of underlying statistical assumptions. Nevertheless, it has been postulated that females exhibit less NREM sleep than males as a result of a lower physiological need for sleep under baseline conditions[57]. This hypothesis is supported by results that indicate shorter NREM sleep time in female mice is not compensated for by higher delta power (a marker of homeostatic sleep drive) during NREM sleep, as NREM delta density is similar between males and females under baseline conditions[57].

We also found that female mice had shorter NREM sleep bouts during the light and dark periods, and longer WAKE bouts during the dark period compared to male mice. These data agree with previous studies for WAKE bout duration, although in those studies no sex differences in NREM sleep bouts were apparent[57,58]. One potential explanation for sex differences in NREM sleep bout duration and the previously reported findings[57,58] could be differences in sample sizes and data collection methods. Notably, our sample sizes (males $n = 140$; females $n = 127$)

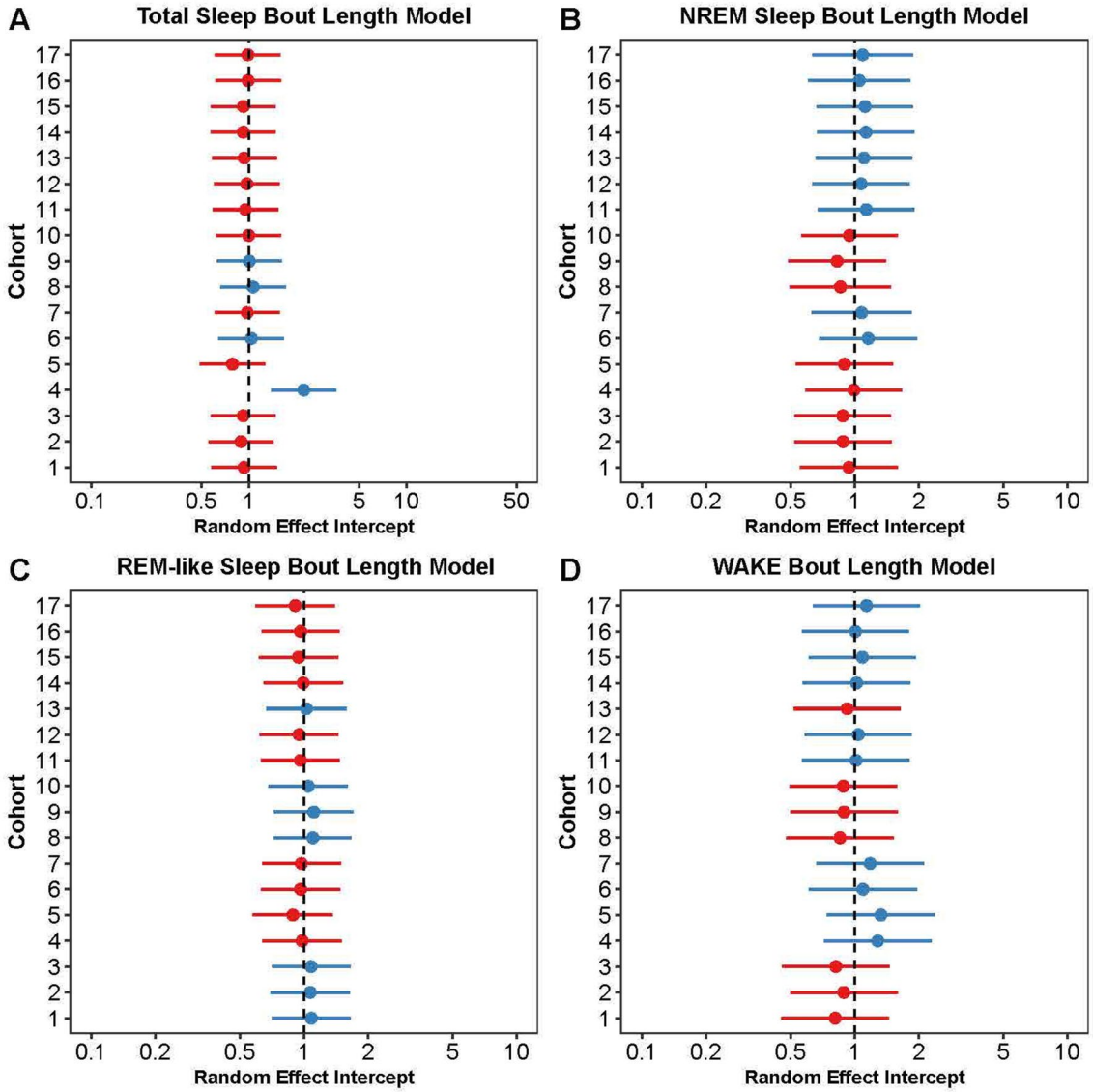

Fig. 6. Total sleep bout lengths nominally varied among cohorts. Estimated cohort-specific random intercepts for (**A**) Total, (**B**) REM, (**C**) Non-REM, and (**D**) WAKE sleep bout lengths. Results shown are point estimates and 95% confidence intervals estimated by hierarchical models with negative-binomial error distributions and random intercepts for individual animals nested within cohorts; blue values denote positive effects, whereas red values denote negative effects.

were 8–10 times greater than those of previous studies, which increased statistical power to detect and accurately characterize true effects, whereas previous studies were likely severely underpowered with high Type II error rates (Supplementary Information S1, Fig. S1.1). The non-invasive piezoelectric recording system- employed herein- allows data collection from a large number of animals in a relatively short period of time because there is no need for invasive EEG/EMG electrode implantation and surgical recovery, which necessarily restricts sample sizes. Thus, a strength of our study, and the non-invasive piezoelectric system, is that the physiological parameters to determine sleep were collected from a very large sample of individuals in the absence of invasive surgical procedures. Based on findings from our study, and reports from others[58], sex differences in the amount of time spent sleeping, and the duration of bout lengths are greatest during the dark period (i.e., active phase). The dark period is when most behaviors, such as eating and reproducing, typically occur; thus, additional wakefulness and shorter sleep bout lengths in females during this active phase may have a selective advantage for the species[58].

The hypothesis that sex differences in the distribution of sleep–wake behavior may be adaptive is supported by observations in *Drosophila*, a model routinely used to study the molecular and circuit basis of sleep regulation, wherein almost all lines of female flies (9000 mutant lines tested) sleep less than males[59]. Specifically, female flies sleep less than males after mating, likely to meet demands of feeding and oviposition[60,61]. Among mammals, female black tufted ear marmosets with young sleep significantly less (are awake more than three times as often) than females without young[62]. Although these observations further support the idea that sexual dimorphisms in sleep could reflect selective advantages for the species, little research has been conducted to test this hypothesis and, as such, this idea remains somewhat speculative.

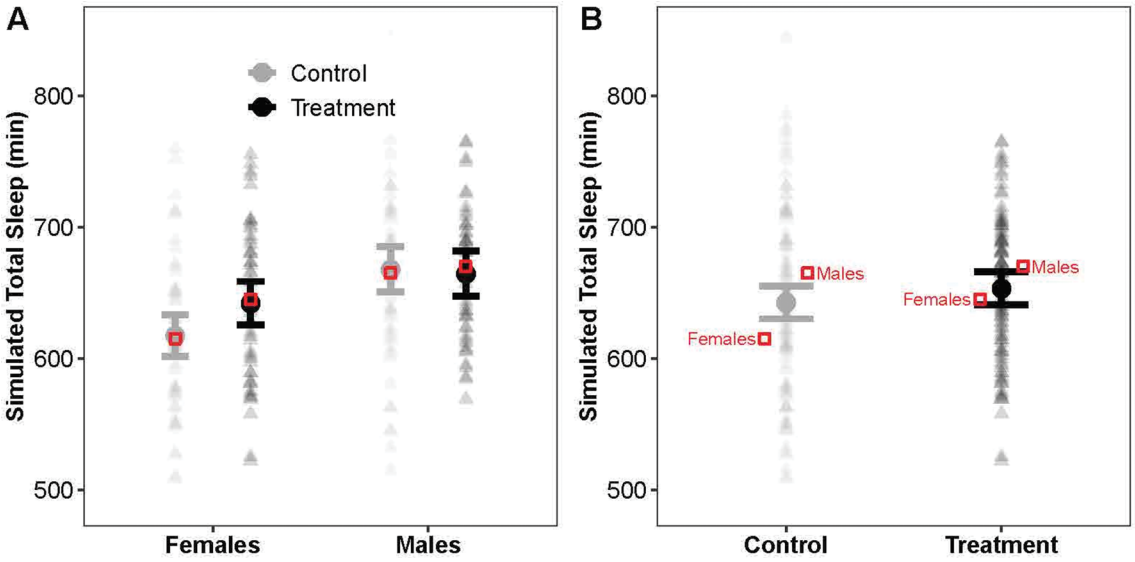

Fig. 7. Simulation demonstrates that mixed sex studies with equal sample sizes may fail to detect true treatment effects. (**A**) Disaggregating by sex and modeling a sex × treatment interaction on simulated sleep data from a hypothetical control versus treatment study accurately estimates the true sex differences and female-specific treatment effect. (**B**) In contrast, conducting a 'mixed sex' analysis in which only a treatment effect on the same simulated sleep data is modeled results in a failure to detect a true sex-specific treatment effect. Results presented as estimated marginal mean point estimates (circles) and their 95% confidence intervals (bars) from generalized linear models with negative-binomial error distributions; true means of the simulated data are depicted by the red squares; background triangles represent the simulated data points for $n = 200$ total hypothetical individuals to which the models were fitted.

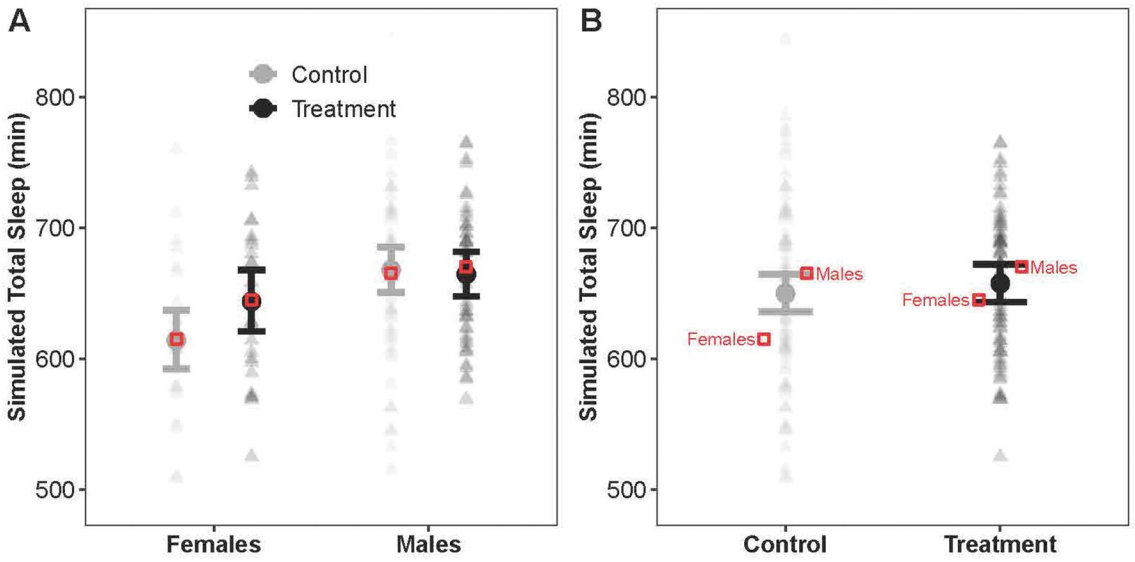

Fig. 8. Simulation demonstrates that unequal sample sizes in mixed sex studies may fail to detect true sex and treatment effects. (**A**) Disaggregating by sex and modeling a sex × treatment interaction on simulated sleep data from a hypothetical control versus treatment study where there were unequal sample sizes accurately estimates the true means but failed to detect the true treatment effect. (**B**) Conducting a 'mixed sex' analysis in which only a treatment effect on the same simulated sleep data is modeled results in a failure to detect a true treatment effect. Results presented as estimated marginal mean point estimates (circles) and their 95% confidence intervals (bars) from generalized linear models with negative-binomial error distributions; true means of the simulated data are depicted by the red squares; background triangles represent the simulated data points for $n = 150$ (females $n = 50$; males $n = 100$) hypothetical individuals to which the models were fitted.

Potential mechanisms driving sexual dimorphisms in sleep include the role of sex hormones on sleep architecture. Evidence exists that a gonadectomy, when performed in male and female C57BL/6 mice, substantially reduces or eliminates sex differences in sleep architecture[58], supporting the hypothesis that gonadal function may drive sex differences in sleep. Importantly, in female mice, hormone replacement returns sleep to baseline levels in gonadectomized mice indicating female sex hormones contribute to sleep quantity[57]. Physiological levels of estradiol alone, or in conjunction with progesterone, modulate sleep architecture of ovariectomized female rats[63]. Similarly, sex differences in human sleep phenotypes exist[18]. Gonadal hormones are hypothesized to mediate sex differences in women, where periods of hormonal fluctuations are associated with an increased prevalence of EEG-measured sleep disturbances[18]. In humans, both sleep duration and quality fluctuate with the menstrual cycle with lower sleep quality reported during the nadir of estradiol and progesterone levels[64,65]. Changes in sleep architecture have also been reported during pregnancy and menopause[64], additional periods of hormonal fluctuations in women.

Major concerns about the replicability of pre-clinical and clinical biomedical research have led funding agencies and journal editors to heighten the transparency of the data collection process. In the current study, mice were individually housed in piezoelectric sleep cages and data were collected across 17 cohorts (~ 16 mice per cohort), with each cohort subjected to sleep monitoring within a different two-day period, to achieve the final sample size of 267 mice. An additional strength of our study is that we explicitly assessed all sleep outcome measures among cohorts, found variation to be generally minimal across them, but directly accounted for all among-cohort variation in our hierarchical statistical models. This lack of variability across cohorts highlights the consistency of the methods used in our laboratory and demonstrates that results from the current study are replicable and reproducible.

We also sought to address replicability issues associated with sample adequacy (i.e., appropriateness of sample size and composition)[66]. Sample size reporting is often poor in biomedical research[66] and, in mixed-sex studies, sex-specific sample sizes are often not reported. Previous rodent studies that investigated sex-dependent dimorphisms in sleep outcomes had sample sizes that were substantially smaller than those reported herein. Based on the observed distribution of total minutes slept from our study, we simulated data that reflected a true sex effect and a hypothetically true female-specific treatment effect. Our simulation exercise showed that when sample sizes were equal: (1) true sex differences and treatment effects were accurately estimated by an appropriately specified model if data were disaggregated by sex, but (2) a female-specific treatment effect was lost if data were not disaggregated by sex. However, for unequal sample sizes, where females were underrepresented, we found that: (1) when data were disaggregated by sex, both the sex differences and the treatment effect were not accurately estimated by an appropriately specified model because of wider confidence intervals caused by the increased uncertainty due to the smaller sample size of females, and (2) female-specific treatment effects were not detected when data were aggregated between sexes. These results highlight the importance of disaggregating sleep data by sex and fitting correctly specified statistical models if a sex-specific effect of a disease/intervention/treatment is possible. These data also emphasize the need for a balanced study design with large sample sizes that are approximately equal between sexes.

Although non-invasive recording techniques have several advantages for sleep research, there are limitations. The piezoelectric non-invasive approach of sleep recording was previously reported using a 2-state decision tree algorithm to discriminate between total sleep (NREM and REM) and wake states with 95% agreement compared to human-scored EEG recordings[23,67,68]. In contrast, we employed a 3-state decision tree algorithm designed to detect all three rodent vigilance states (Sleep Stats Version 4; Signal Solutions, Lexington, KY, USA). This 3-state decision tree algorithm was successfully validated in mice[25] and rats[26], and further validation in mice is presented in Supplementary Information S2. For total sleep (combining NREM and REM) and wake, agreement with EEG/EMG was in the 90% range for both precision and recall for our data obtained using Sleep Stats Version 4. However, when comparing the occurrence of REM on 2-second intervals with the EEG/EMG scored results, the REM-like classification drops to a 50% agreement in large part because the non-invasive piezo system measures behaviors, such as activity and breathing (thorax motion), that are correlated with the brain state but cannot directly measure the brain state (see Supplementary Information S2, Table S2.2). To distinguish REM-like sleep, the algorithm detects characteristic increases in respiration rate and variability in signal amplitude that do not typically occur during NREM sleep. While the REM brain state does increase the likelihood of these irregularities during sleep, there is not a strong correlation to the onset and offset times of the event. In other words, the piezo system relies on the measurement of subtle physical manifestations correlating to the REM state but cannot directly measure REM sleep as it does not measure brain activity. For this reason, the authors have chosen to report REM data as 'REM-like' sleep.

As a result, agreement between piezo-scored REM and EEG-scored REM is best when looking at sleep percentages over larger time intervals.

Classification of REM-like sleep by non-EEG/EMG assessment methods can be difficult as increased autonomic activity increases variability in parameters used to define sleep states. For example, determination of REM-like sleep using video-based methods is much less accurate than for NREM sleep or wakefulness (WAKE)[69]. Nevertheless, and although there are limitations to the piezo system, this approach has utility for studies that may be difficult to do using conventional time- and labor-intensive EEG/EMG recording, which are also fraught with inaccuracies (e.g., agreement between human scorers on EEG/EMG was ~ 84%; Supplementary Information S2). For example, surgical implantation of EEG/EMG recording electrodes into the sample size of mice required to achieve 80% power for detecting a 5% difference in NREM sleep between groups ($n = 240$ individuals), and then analyzing data using standard methods, would require an incredible investment of time, resources, and effort. Moreover, EEG electrode implantation causes inflammation[70], which can confound the results of studies investigating pathological-induced sleep changes, further highlighting the importance of non-invasive options for measuring sleep in the rodent. The non-invasive system used in our study is a reliable alternative to invasive EEG

recordings and allows for data to be collected from large samples through experimental procedures that minimize or eliminate pain or distress in rodents. Importantly, our reported results for NREM sleep, REM-like sleep, and WAKE for male mice are similar to published results from other studies that used EEG to determine sleep in male C57BL/6 mice[71–75]. A direct comparison of parameters determined from sleep studies of male C57BL/6 mice during baseline conditions is provided in the Supplementary Information S2 (Supplementary Table S2.1).

An additional consideration is that, in order to minimize handling and non-specific stress responses, we did not track the estrous cycles of female mice. As previously mentioned, clinical research has shown that female sleep architecture changes with fluctuating levels of endogenous estradiol and progesterone, and in response to hormonal treatments (e.g., oral contraceptives, hormone replacement)[63,76]. Similarly, rodent studies have shown that sleep architecture in females changes as a function of their hormonal milieu. In the absence of estrous cycle monitoring, we cannot determine the extent to which dimorphic sex effects on sleep architecture in the current study are hormone-driven, and future studies are warranted to investigate the role of sex hormones on sleep regulation.

In conclusion, we observed profound biological sex differences in sleep with female mice exhibiting more wakefulness (WAKE) and less NREM sleep with shorter NREM bout lengths. Here, we also outline research implications for this underlying biological sex difference and recommend how data should be interpreted. Although there are sex differences in the sleep of humans[77], far less is understood about sex differences in the sleep of rodents, a common model used in biomedical research. Despite the critical role sleep has in various health outcomes, few studies account for underlying sex effects in analyses of rodent sleep. As sleep increasingly becomes a focus of biomedical sciences, it is particularly important to avoid reporting confounded findings. We recommend that sleep studies using male and female mice consider underlying biological sex differences in sleep outcomes when analyzing data and interpreting results. Through simulations, we demonstrated the importance of disaggregating data in statistical analyses to determine if sex-specific effects exist. We also show that it may be critical to balance study designs with equal representation and adequately large sample sizes of both sexes. We further demonstrate the feasibility of using a non-invasive method as a reliable alternative to invasive EEG recordings in the mouse. Finally, additional research is needed to elucidate the cellular and molecular mechanisms underlying the observed biological sex differences reported in this study.

Data availability

Data are publicly available in the Dryad Digital Repository: https://doi.org/10.5061/dryad.76hdr7t39

Received: 15 April 2024; Accepted: 22 August 2024
Published online: 15 October 2024

References

1. Farhud, D. & Aryan, Z. Circadian rhythm, lifestyle and health: A narrative review. *Iran J. Public Health* **47**, 1068–1076 (2018).
2. Abbott, S. M. & Videnovic, A. Chronic sleep disturbance and neural injury: Links to neurodegenerative disease. *Nat. Sci. Sleep* **8**, 55–61. https://doi.org/10.2147/NSS.S78947 (2016).
3. Wolk, R., Gami, A. S., Garcia-Touchard, A. & Somers, V. K. Sleep and cardiovascular disease. *Curr. Probl. Cardiol.* **30**, 625–662. https://doi.org/10.1016/j.cpcardiol.2005.07.002 (2005).
4. Yin, D., Chen, S. & Liu, J. Sleep disturbances in autoimmune neurologic diseases: Manifestation and pathophysiology. *Front. Neurosci.* **15**, 687536. https://doi.org/10.3389/fnins.2021.687536 (2021).
5. van Leeuwen, W. M. *et al.* Sleep restriction increases the risk of developing cardiovascular diseases by augmenting proinflammatory responses through IL-17 and CRP. *PloS one* **4**, e4589. https://doi.org/10.1371/journal.pone.0004589 (2009).
6. Medic, G., Wille, M. & Hemels, M. E. Short- and long-term health consequences of sleep disruption. *Nat. Sci. Sleep* **9**, 151–161. https://doi.org/10.2147/NSS.S134864 (2017).
7. Lv, Y. N., Cui, Y., Zhang, B. & Huang, S. M. Sleep deficiency promotes Alzheimer's disease development and progression. *Front. Neurol.* **13**, 1053942. https://doi.org/10.3389/fneur.2022.1053942 (2022).
8. Worley, S. L. The extraordinary importance of sleep: The detrimental effects of inadequate sleep on health and public safety drive an explosion of sleep research. *P. T.* **43**, 758–763 (2018).
9. Morssinkhof, M. W. L. *et al.* Associations between sex hormones, sleep problems and depression: A systematic review. *Neurosci. Biobehav. Rev.* **118**, 669–680. https://doi.org/10.1016/j.neubiorev.2020.08.006 (2020).
10. Bierer, B. E., Meloney, L. G., Ahmed, H. R. & White, S. A. Advancing the inclusion of underrepresented women in clinical research. *Cell Rep. Med.* **3**, 100553. https://doi.org/10.1016/j.xcrm.2022.100553 (2022).
11. Uhl, K., Parekh, A. & Kweder, S. Females in clinical studies: Where are we going?. *Clin. Pharmacol. Ther.* **81**, 600–602. https://doi.org/10.1038/sj.clpt.6100112 (2007).
12. Vidaver, R. M., Lafleur, B., Tong, C., Bradshaw, R. & Marts, S. A. Women subjects in NIH-funded clinical research literature: Lack of progress in both representation and analysis by sex. *J. Womens Health Gend. Based Med.* **9**, 495–504. https://doi.org/10.1089/15246090050073576 (2000).
13. Arnegard, M. E., Whitten, L. A., Hunter, C. & Clayton, J. A. Sex as a biological variable: A 5-year progress report and call to action. *J. Womens Health (Larchmt)* **29**, 858–864. https://doi.org/10.1089/jwh.2019.8247 (2020).
14. Mong, J. A. & Cusmano, D. M. Sex differences in sleep: Impact of biological sex and sex steroids. *Philos. Trans. R. Soc. B: Biol. Sci.* **371**, 20150110. https://doi.org/10.1098/rstb.2015.0110 (2016).
15. Mallampalli, M. P. & Carter, C. L. Exploring sex and gender differences in sleep health: A society for women's health research report. *J. Women's Health* **23**, 553–562. https://doi.org/10.1089/jwh.2014.4816 (2014).
16. Zendels, P., Ruggiero, A. & Gaultney, J. Gender differences affecting the relationship between sleep attitudes, sleep behaviors and sleep outcomes. *Cogent Psychol.* **8**, 1 (2021).
17. Bryant, C. D. The blessings and curses of C57BL/6 substrains in mouse genetic studies. *Ann. N. Y. Acad. Sci.* **1245**, 31–33. https://doi.org/10.1111/j.1749-6632.2011.06325.x (2011).
18. Dib, R., Gervais, N. J. & Mongrain, V. A review of the current state of knowledge on sex differences in sleep and circadian phenotypes in rodents. *Neurobiol. Sleep Circadian Rhythms* **11**, 100068. https://doi.org/10.1016/j.nbscr.2021.100068 (2021).
19. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412. https://doi.org/10.1371/journal.pbio.1000412 (2010).

20. Rowe, R. K. *et al.* Acute post-traumatic sleep may define vulnerability to a second traumatic brain injury in mice. *J. Neurotrauma* https://doi.org/10.1089/neu.2018.5980 (2018).
21. Harrison, J. L. *et al.* Resolvins AT-D1 and E1 differentially impact functional outcome, post-traumatic sleep, and microglial activation following diffuse brain injury in the mouse. *Brain Behav. Immun.* **47**, 131–140. https://doi.org/10.1016/j.bbi.2015.01.001 (2015).
22. Saber, M. *et al.* Acute peripheral inflammation and post-traumatic sleep differ between sexes after experimental diffuse brain injury. *Eur. J. Neurosci.* https://doi.org/10.1111/ejn.14611 (2019).
23. Donohue, K. D., Medonza, D. C., Crane, E. R. & O'Hara, B. F. Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice. *Biomed. Eng. Online* **7**, 14 (2008).
24. Ouellette, G. & Donohue, K. D. *Sleep Stats Data Explorer*. Vol. 4 (Signal Solutions, LLC, 2022).
25. Yaghouby, F., Donohue, K. D., O'Hara, B. F. & Sunderam, S. Noninvasive dissection of mouse sleep using a piezoelectric motion sensor. *J. Neurosci. Methods* **259**, 90–100. https://doi.org/10.1016/j.jneumeth.2015.11.004 (2016).
26. Topchiy, I., Fink, A. M., Maki, K. A. & Calik, M. W. Validation of PiezoSleep scoring against EEG/EMG sleep scoring in rats. *Nat. Sci. Sleep* **14**, 1877–1886. https://doi.org/10.2147/NSS.S381367 (2022).
27. Faraway, J. J. *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models* (Chapman and Hall, 2016).
28. Stroup, W. W. *Generalized linear mixed models: Modern concepts, methods, and applications* (CRC Press, 2012).
29. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R. J.* **9**, 378–400. https://doi.org/10.32614/Rj-2017-066 (2017).
30. Team, R. C. *R: A language and environment for statistical computing.*, (2022).
31. Hilbe, J. M. Modeling count data. *Modeling Count Data*, 1–283, https://doi.org/10.1017/Cbo9781139236065 (2014).
32. Ferrari, S. & Cribari-Neto, F. Beta regrssion for modelling rates and proportions. *Appl. Stat.* **31**, 799–815 (2004).
33. Kedem, B. F. *Regression models for time series analysis* (John Wiley & Sons, 2002).
34. Woolridge, J. M. in *Introductory econometrics: A modern approach, Fifth Edition* 343–446, South-Western Cenage Learning, 2016.
35. Rowe, R. K. *et al.* Microglia are necessary to regulate sleep after an immune challenge. *Biology* **11**, 1241. https://doi.org/10.3390/biology11081241 (2022).
36. Saber, M. *et al.* Acute peripheral inflammation and post-traumatic sleep differ between sexes after experimental diffuse brain injury. *Eur. J. Neurosci.* **52**, 2791–2814. https://doi.org/10.1111/ejn.14611 (2020).
37. Saber, M., Murphy, S. M., Cho, Y., Lifshitz, J. & Rowe, R. K. Experimental diffuse brain injury and a model of Alzheimer's disease exhibit disease-specific changes in sleep and incongruous peripheral inflammation. *J. Neurosci. Res.* **99**, 1136–1160. https://doi.org/10.1002/jnr.24771 (2021).
38. Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. & Schmid, M. A review of spline function procedures in R. *Bmc Med. Res. Methodol.* **19**, 46. https://doi.org/10.1186/s12874-019-0666-3 (2019).
39. de Boor, C. Calculation of the smoothing spline with weighted roughness measure. *Math. Models Methods Appl. Sci.* **11**, 33–41 (2001).
40. Wang, Y. *Smoothing splines: Methods and applications* (Chapman and Hall/CRC, 2011).
41. Burnham, K. P., Anderson, D. R. & Huyvaert, K. P. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65**, 23–35 (2011).
42. Burnham, K. P. & Anderson, D. R. *Model selection and multimodel inference: A practical information-theoretic approach* (Springer, 2002).
43. Stroup, W. W. *Generalized linear mixed models: Modern concepts, methods and applications* (CRC Press, 2012).
44. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).
45. Sawilowsky, S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **8**, 467–474 (2009).
46. Lenth, R. V. B., P.; Herve, M.; Love, J.; Miguez, F.; Riebl, H.; Singmann, H. . *emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.3*, 2022).
47. Sullivan, G. M. & Feinn, R. Using effect size-or why the p value is not enough. *J. Grad. Med. Educ.* **4**, 279–282. https://doi.org/10.4300/JGME-D-12-00156.1 (2012).
48. Morris, T. P., White, I. R. & Crowther, M. J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **38**, 2074–2102. https://doi.org/10.1002/sim.8086 (2019).
49. Venables, W. N. & Ripley, B. D. *Modern applied statistics with S* 4th edn. (Springer, 2002).
50. Chalmers, R. P. & Adkins, M. C. Writing effective and reliable Monte Carlo simulations with the SimDesign package. *Quant Meth. Psychol.* **16**, 248–280. https://doi.org/10.20982/tqmp.16.4.p248 (2020).
51. Liu, Y. *et al.* Association between perceived insufficient sleep, frequent mental distress, obesity and chronic diseases among US adults, 2009 behavioral risk factor surveillance system. *BMC Public Health* **13**, 84. https://doi.org/10.1186/1471-2458-13-84 (2013).
52. Banks, S. & Dinges, D. F. Behavioral and physiological consequences of sleep restriction. *J. Clin. Sleep Med.: JCSM : Off. Publ. Am. Acad. Sleep Med.* **3**, 519–528 (2007).
53. Reimer, M. A. & Flemons, W. W. Quality of life in sleep disorders. *Sleep Med. Rev.* **7**, 335–349. https://doi.org/10.1053/smrv.2001.0220 (2003).
54. Sigurdson, K. & Ayas, N. T. The public health and safety consequences of sleep disorders. *Can. J. Physiol. Pharmacol.* **85**, 179–183. https://doi.org/10.1139/y06-095 (2007).
55. Green, T. R. F., Ortiz, J. B., Wonnacott, S., Williams, R. J. & Rowe, R. K. The bidirectional relationship between sleep and inflammation links traumatic brain injury and alzheimer's disease. *Front. Neurosci.* **14**, 894. https://doi.org/10.3389/fnins.2020.00894 (2020).
56. Rowe, R. K. & Griesbach, G. S. Immune-endocrine interactions in the pathophysiology of sleep-wake disturbances following traumatic brain injury: A narrative review. *Brain Res. Bull.* **185**, 117–128. https://doi.org/10.1016/j.brainresbull.2022.04.017 (2022).
57. Choi, J. *et al.* The role of reproductive hormones in sex differences in sleep homeostasis and arousal response in mice. *Front. Neurosci.* **15**, 739236. https://doi.org/10.3389/fnins.2021.739236 (2021).
58. Paul, K. N., Dugovic, C., Turek, F. W. & Laposky, A. D. Diurnal sex differences in the sleep-wake cycle of mice are dependent on gonadal function. *Sleep* **29**, 1211–1223. https://doi.org/10.1093/sleep/29.9.1211 (2006).
59. Cirelli, C. *et al.* Reduced sleep in Drosophila Shaker mutants. *Nature* **434**, 1087–1092. https://doi.org/10.1038/nature03486 (2005).
60. Isaac, R. E., Li, C., Leedale, A. E. & Shirras, A. D. Drosophila male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. *Proc. Biol. Sci.* **277**, 65–70. https://doi.org/10.1098/rspb.2009.1236 (2010).
61. Asahina, K. Sex differences in drosophila behavior: qualitative and quantitative dimorphism. *Curr. Opin. Physiol.* **6**, 35–45. https://doi.org/10.1016/j.cophys.2018.04.004 (2018).
62. Fite, J. E. *et al.* Nighttime wakefulness associated with infant rearing in. *Int. J. Primatol.* **24**, 1267–1280 (2003).
63. Deurveilher, S., Rusak, B. & Semba, K. Female reproductive hormones alter sleep architecture in ovariectomized rats. *Sleep* **34**, 519–530. https://doi.org/10.1093/sleep/34.4.519 (2011).
64. Harrington, Y. A. *et al.* Sex hormones, sleep, and memory: interrelationships across the adult female lifespan. *Front. Aging Neurosci.* **14**, 800278. https://doi.org/10.3389/fnagi.2022.800278 (2022).
65. Baker, F. C. & Driver, H. S. Self-reported sleep across the menstrual cycle in young, healthy women. *J. Psychosom. Res.* **56**, 239–243. https://doi.org/10.1016/S0022-3999(03)00067-9 (2004).

66. Vasileiou, K., Barnett, J., Thorpe, S. & Young, T. Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *Bmc Med. Res. Methodol.* **18**, 148. https://doi.org/10.1186/s12874-018-0594-7 (2018).
67. Mang, G. M. *et al.* Evaluation of a piezoelectric system as an alternative to electroencephalogram/electromyogram recordings in mouse sleep studies. *Sleep* **37**, 1383–1392. https://doi.org/10.5665/sleep.3936 (2014).
68. Flores, A. E. *et al.* Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements. *IEEE Trans. Bio-med. Eng.* **54**, 225–233 (2007).
69. McShane, B. B. *et al.* Assessing REM sleep in mice using video data. *Sleep* **35**, 433–442 (2012).
70. Balzekas, I., Hernandez, J., White, J. & Koh, S. Confounding effect of EEG implantation surgery: Inadequacy of surgical control in a two hit model of temporal lobe epilepsy. *Neurosci. Lett.* **622**, 30–36. https://doi.org/10.1016/j.neulet.2016.04.033 (2016).
71. Franken, P., Malafosse, A. & Tafti, M. Genetic determinants of sleep regulation in inbred mice. *Sleep* **22**, 155–169 (1999).
72. Deboer, T., Ruijgrok, G. & Meijer, J. H. Short light-dark cycles affect sleep in mice. *Eur. J. Neurosci.* **26**, 3518–3523. https://doi.org/10.1111/j.1460-9568.2007.05964.x (2007).
73. Veasey, S. C. *et al.* An automated system for recording and analysis of sleep in mice. *Sleep* **23**, 1025–1040 (2000).
74. Toth, L. A. & Opp, M. R. Cytokine- and microbially induced sleep responses of interleukin-10 deficient mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **280**, 1806–1814 (2001).
75. Morrow, J. D. & Opp, M. R. Diurnal variation of lipopolysaccharide-induced alterations in sleep and body temperature of interleukin-6-deficient mice. *Brain Behav. Immun.* **19**, 40–51. https://doi.org/10.1016/j.bbi.2004.04.001 (2005).
76. Shechter, A. & Boivin, D. B. Sleep, hormones, and circadian rhythms throughout the menstrual cycle in healthy women and women with premenstrual dysphoric disorder. *Int. J. Endocrinol.* **2010**, 259345. https://doi.org/10.1155/2010/259345 (2010).
77. Carrier, J. *et al.* Sex differences in age-related changes in the sleep-wake cycle. *Front. Neuroendocrinol.* **47**, 66–85. https://doi.org/10.1016/j.yfrne.2017.07.004 (2017).

Acknowledgements

We acknowledge Lindsey Beauregard, Madison Szewczak, and Shiying Zou for technical assistance.

Author contributions

G.S.M. assisted with experimental design, data collection, and led writing of the manuscript. T.R.F.G. assisted with experimental design, data collection, and reviewed and edited the manuscript. S.M.M. led statistical analyses and visualization, and reviewed and edited the manuscript. M.R.O. assisted with conceptualization and reviewed and edited the manuscript. R.K.R. conceptualized the study, developed the experimental design, assisted with the formal analysis, and reviewed and edited the manuscript. K.D.D. generated validation data presented in the Supplementary Information.

Funding

RKR, MRO, and TRFG were supported, in part, by funds from the Department of Defense Congressionally Directed Medical Research Programs (PR211450, PR211450P1). GSM was supported by the University of Colorado Boulder Undergraduate Research Opportunities Program (UROP) fellowship.

Competing interests

KDD is a partial owner of Signal Solution LLC and provided supplementary data for the validation of the three-state algorithm. No other aspects of the research presented here were sponsored by Signal Solution. All the remaining authors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-70996-1.

Correspondence and requests for materials should be addressed to R.K.R.