Received: 27 October 2021 | Revised: 27 January 2022 | Accepted: 7 March 2022
DOI: 10.1002/sim.9385

RESEARCH ARTICLE

Statistics in Medicine WILEY

Jointly modeling of sleep variables that are objectively measured by wrist actigraphy

Xiaonan Xue[1] | **Simin Hua**[1] | **Kathleen Weber**[2] | **Qibin Qi**[1] | **Robert Kaplan**[1] | **Deborah R. Gustafson**[3] | **Anjali Sharma**[4] | **Audrey French**[2,5] | **Helen J. Burgess**[6]

[1]Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, New York, USA

[2]Core Center of Cook County Health, Stroger Hospital of Cook County, Chicago, Illinois, USA

[3]Department of Neurology, Section for Neuro Epidemiology, State University of New York Downstate Health Sciences University, Brooklyn, New York, USA

[4]Department of Medicine, Albert Einstein College of Medicine, Bronx, New York, USA

[5]Division of Infectious Diseases, Stroger Hospital of Cook County, Chicago, Illinois, USA

[6]Sleep and Circadian Research Laboratory, Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Xiaonan Xue, Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA.
Email: xiaonan.xue@einsteinmed.org

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: 1R01HL142116-01

Abstract

Recently developed actigraphy devices have made it possible for continuous and objective monitoring of sleep over multiple nights. Sleep variables captured by wrist actigraphy devices include sleep onset, sleep end, total sleep time, wake time after sleep onset, number of awakenings, etc. Currently available statistical methods to analyze such actigraphy data have limitations. First, averages over multiple nights are used to summarize sleep activities, ignoring variability over multiple nights from the same subject. Second, sleep variables are often analyzed independently. However, sleep variables tend to be correlated with each other. For example, how long a subject sleeps at night can be correlated with how long and how frequent he/she wakes up during that night. It is important to understand these inter-relationships. We therefore propose a joint mixed effect model on total sleep time, number of awakenings, and wake time. We develop an estimating procedure based upon a sequence of generalized linear mixed effects models, which can be implemented using existing software. The use of these models not only avoids computational intensity and instability that may occur by directly applying a numerical algorithm on a complicated joint likelihood function, but also provides additional insights on sleep activities. We demonstrated in simulation studies that the proposed estimating procedure performed well in estimating both fixed and random effects' parameters. We applied the proposed model to data from the Women's Interagency HIV Sleep Study to examine the association of employment status and age with overall sleep quality assessed by several actigraphy measured sleep variables.

KEYWORDS

compound Poisson gamma distribution, generalized linear mixed effects model, Poisson distribution with over-dispersion, Tweedie distribution

1 | INTRODUCTION

Over recent years, there has been considerable interest in examining the relationship between sleep and health outcomes. Traditionally, self-reported sleep information is collected via questionnaires. Short sleep duration measured subjectively has been shown to be associated with increased BMI,[1] impaired glucose tolerance,[2] diabetes,[3,4] and all-cause mortality.[5] The availability of wrist actigraphy allows an objective and continuous assessment of sleep parameters[6-8] and permits multiple nights of recording at home with minimal participant burden.[6,9,10] Compared to objectively measured sleep duration, subjectively measured sleep duration is less accurate and therefore tends to bias the association between sleep duration and health outcomes towards the null. For example, a study on individuals with insomnia showed that objectively measured but not subjectively measured short sleep duration increased their risk for hypertension.[11] In this study, using the mean of 2 nights of polysomnography measured total sleep time as the gold standard, the mean of total sleep time based on 2 weeks of subjective sleep diary reports had 60% of sensitivity and 64% specificity in detecting short sleep duration, that is, total sleep time less than 6 hours. Because of the low inaccuracy in the subjectively measured total sleep time, the area under the ROC curve for hypertension was only 0.639 using the subjectively measured total sleep time. Furthermore, actigraphy can be used to continuously measure sleep and wake patterns at night. Sleep and wake patterns measured by actigraphy showed that decreased sleep efficiency and increased sleep fragmentation are associated with hypertension.[12] Actigraphic estimates of sleep-related problems including wake after sleep onset and number of awakenings were shown to be higher among HIV-infected children than HIV-uninfected children.[13]

Our work was motivated by the Women's Interagency HIV Sleep Study, an ongoing longitudinal study started in 2018 to examine the relationship between sleep, circadian disruption and the tryptophan-kynurenine (TRP-KYN) pathway. The study enrolled over 300 women primarily from the Women's Interagency HIV Study (WIHS), a multi-site longitudinal observational study of midlife women living with HIV and demographically similar HIV-uninfected women with history of risk of HIV.[14] A small number of additional women at the Chicago site of WIHS were recruited from the same clinical care setting as WIHS women. The study population is predominantly low-income urban dwelling women of color. Participants were asked to wear a wrist actigraphy monitor (Actiwatch Spectrum Plus, 30-second epochs) for multiple nights (details of actigraphy data extraction and interpretation are given in Section 4). Objective actigraphic estimates of sleep timing (sleep onset time, sleep end time) and total sleep time were extracted. Sleep onset is the clock time of the first epoch scored as sleep during the rest interval; sleep end is the clock time of the last epoch scored as sleep, that is, the final wake time, during the rest interval. Total sleep time is the sum of all epochs scored as sleep between sleep onset and sleep end. Measures of sleep continuity included sleep efficiency, wake after sleep onset (WASO), and the number of wake bouts were also extracted. Sleep efficiency is defined as the proportion of time between sleep onset and sleep end time scored as sleep in the rest interval, expressed as a percentage. WASO is defined as the sum of all epochs scored as wake within the rest interval. Wake bout is defined as the continuous period of wake, at least 30 seconds long during the rest interval. Because the actigraphy collects data every 30 seconds, the number of wake bouts equals to the number of awakenings.

Sleep variables are typically averaged over multiple nights to summarize a subject's sleep activites,[15] ignoring repeated measures from the same subject. Furthermore, sleep variables are often analyzed separately.[16,17] However, a single sleep variable does not provide a full assessment of sleep. In addition, sleep variables tend to be correlated with each other. For example, total sleep time can be related to how many times and how long a subject wakes up at night, and how frequent and how long a subject wakes up at night may also be related to each other. In this article, we propose to analyze all repeated observations from the same subject together and to analyze total sleep time, number of awakenings, and WASO simultaneously. We aim to gain insight on how risk factors affect the overall quality of sleep measured by several sleep variables and how improvement in one sleep variable may affect the others.

Sleep data are somewhat similar to data on use of healthcare resources, where health insurance claims and costs during a given insured period are modeled and analyzed.[18-20] Total healthcare cost is usually right skewed and semi-continuous with a point mass at zero for non-users. Therefore, common models for continuous data such as a gamma distribution or a log-normal distribution are not appropriate. A zero-inflated model is not appropriate either because the underlying assumption of two-stage decision process, that is, use or not use healthcare and how much to use are not usually made.[21] Therefore, health economics models often assume the number of healthcare claims follows a Poisson distribution and if there is any claim, the cost of each claim follow a gamma distribution and they are assumed to be independent with each other.[21] Thus, the total healthcare cost follows a compound Poisson gamma distribution, a distribution on semi-continuous data that allows exact zero. A compound Poisson gamma distribution does not have a closed form marginal distribution. It belongs to the family of Tweedie distributions.[20-23] In the context of sleep data, the number of health insurance claims in a given period corresponds to the number of awakenings at a night and the cost of each health insurance claim corresponds to the duration of each wake bout thus the total cost of healthcare in the given period corresponds to the total wake time, that is, WASO, at that night.

Despite the similarity between healthcare and sleep data, there are major differences between them. First, in healthcare data the total number of health insurance claims and the total cost for all claims aggregated over an insured period are collected once for every participant under study. No repeated observations from the same participant are involved. In sleep studies, sleep variables are typically collected for each participant over multiple nights. Second, in healthcare data the duration of the insured period is treated as fixed and it is typically not to be influenced by the number of health care use and the cost per use. In sleep studies the corresponding variable, the total sleep time, is one of the main sleep variables that we aim to model and to examine its relationship with the number of awakenings and the duration per wake episode. Therefore, in this article, we adopt traditional models used for healthcare data such that we assume the number of awakenings follows a Poisson distribution and if there is any wake episode, the duration of each wake bout follow a gamma distribution. But we further extend the model to incorporate repeated observations from the same participant and to model the total sleep time as one of the outcome variables and furthermore to develop a joint model on total sleep time, number of awakening and duration of wake bouts. To our best knowledge, currently there is no statistical method available to jointly model sleep variables that are repeatedly measured over multiple nights.

This article is organized as follows: a joint model for the sleep variables and its estimating procedure are described in Section 2. In Section 3, a simulation study is used to examine the performance of the estimation procedure. The proposed model and methods are applied to the WIHS Sleep Study in Section 4 to examine effects of risk factors on sleep variables and to assess the level of heterogeneity in each sleep variable across participants as well as the inter-relationships among sleep variables. The article is concluded in Section 5.

2 | MODELS AND METHODS

Let T_{ik} represents the total sleep time, M_{ik} represents the number of wake bouts (ie, the number of awakenings), and Y_{ijk} represents the duration of the wake bout at the jth awakening for the ith person on the kth night, where $j = 1, \dots, M_{ik}$ and $k = 1, \dots, K_i$, for example, $K_i = 3$ to 7 nights, $i = 1, \dots, N$. We assume total sleep time follow a gamma distribution. We further assume that the number of awakenings follows a Poisson distribution; and if there is any awakening, the duration of each wake bout also follows a gamma distribution but with different shape and scale parameters than the total sleep time. Random effects are used to incorporate heterogeneity across participants. Specifically, we assume the total sleep time

$$T_{ik} \mid v_{1i} \sim \text{Gamma}\left(\alpha_S, v_{1i} r_i\right), \tag{1}$$

where α_S is the shape parameter and $v_{1i} r_i$ is the scale parameter for subject i and r_i is the fixed effect and v_{1i} is the random effect.

We assume the number of awakening at a night follows a Poisson distribution. How many times a subject wakes up at a night depends on how long this subject sleeps at that night: the longer the subject sleeps at a night, the more awakenings this subject is likely to have at that night. Therefore, we assume

$$M_{ik} \mid T_{ik}, v_{2i} \sim \text{Poisson}\left(v_{2i} \lambda_i T_{ik}\right), \tag{2}$$

where $v_{2i}\lambda_i$ is the unit rate for awakenings for subject i. Here v_{2i} is the random effect and λ_i is the fixed effect for subject i. Next, wake bout durations are assumed to follow a gamma distribution with the same shape parameter α_W across all bouts, that is,

$$Y_{ijk} \mid v_{3i} \sim \text{Gamma}\left(\alpha_W, v_{3i} \xi_i\right), \tag{3}$$

where $v_{3i}\xi_i$ is the scale parameter for subject i. Here v_{3i} is the random effect and ξ_i is the fixed effect for subject i. Conditional on (v_{1i}, v_{2i}, v_{3i}), we assume $M_{ik} \perp Y_{ijk}$ and $Y_{ijk} \perp T_{ik}$.

The total sleep time, the number of awakenings, and wake bout durations may be correlated with each other. It is important to learn, for example, if people who tend to wake up more frequently tend to sleep less and if people who tend to wake up more frequently tend to wake at a shorter duration each episode. Thus, we allow these random effects to be correlated with each other, that is,

$$w_i = (w_{1i}, w_{2i}, w_{3i}) = \left(\log v_{1i}, \log v_{2i}, \log v_{3i}\right) \sim \text{MVN}(0, \Sigma),$$

where $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$.

Below we propose a mixed effect regression model on each sleep variable.

2.1 | A mixed effects model for total sleep time, the number of awakenings and wake time

Conditional on υ_{1i}, we model the expected total sleep time by

$$\log E\left(T_{ik}|\upsilon_{1i}\right) = \gamma_0 + \gamma_1 X_{1i} + w_{1i} \tag{4}$$

so that $E\left(T_{ik}|\upsilon_{1i}\right) = \alpha_S r_i \upsilon_{1i} = e^{\gamma_0+\gamma_1 X_{1i}}\upsilon_{1i}$, where X_{1i} is the covariate vector for the ith person. Since $E\left(T_{ik}\right) = \alpha_S r_i e^{1/2\sigma_1^2}$, e^{γ_1} also represents the conditional as well marginal mean ratio in total sleep time associated with X_1. Conditional on υ_{2i}, we model the expected number of awakenings by

$$\log E\left(M_{ik}|T_{ik}, \upsilon_{2i}\right) = \beta_0 + \beta_1 X_{2i} + \log\left(T_{ik}\right) + w_{2i}, \tag{5}$$

where X_{2i} is the covariate vector for the ith person so that $\lambda_i = e^{\beta_0+\beta_1 X_{2i}}$. As the model is conditioning on the total sleep time, Equation (5) models the frequency of awakenings. Note that $E\left(M_{ik}|T_{ik}, \upsilon_{2i}\right) = \lambda_i T_{ik}\upsilon_{2i}$ so that $E\left(M_{ik}|T_{ik}\right) = \lambda_i T_{ik} e^{1/2\sigma_2^2}$. Therefore, e^{β_1} represents the conditional as well as marginal ratio in rate of awakenings associated with X_2.

Conditional on $M_{ik} > 0$ and υ_{3i}, the distribution of WASO is

$$Y_{i\cdot k} = \sum_{j=1}^{M_{ik}} Y_{ijk} \sim \text{Gamma}\left(M_{ik}\alpha_W, \upsilon_{3i}\xi_i\right).$$

We model the expected WASO by

$$\log E\left(Y_{i\cdot k}|M_{ik}, \upsilon_{3i}\right) = \delta_0 + \delta_1 X_{3i} + \log\left(M_{ik}\right) + w_{3i} \tag{6}$$

so that $E\left(Y_{i\cdot k}|M_{ik}, \upsilon_{3i}\right) = \alpha_W \xi_i M_{ik}\upsilon_{3i} = e^{\delta_0+\delta_1 X_{3i}} M_{ik}\upsilon_{3i}$, where X_{3i} is the covariate vector for the ith person. Since $E\left(Y_{i\cdot k}|M_{ik}\right) = \alpha_W \xi_i M_i e^{1/2\sigma_3^2}$, similarly as above, e^{δ_1} is the conditional as well marginal ratio in the duration of a wake bout associated with X_3. Note that X_1, X_2, and X_3 may or may not be the same. Here for the ease of notation, we assume $X_1 = X_2 = X_3$.

2.2 | Estimation procedure

The joint likelihood function for the total sleep time, the number of awakenings, and WASO is

$$L = \prod_{i=1}^{N} \oint_{W_i} \prod_{k=1}^{K_i} f_T\left(t_{ik}|\gamma_0, \gamma_1, \alpha_S, w_{3i}\right)\left[P_M\left(m_{ik} = 0|\beta_0, \beta_1, t_{ik}, w_{1i}\right)\right]^{I\left(m_{ik}=0\right)}\left[P_M\left(m_{ik}|\beta_0, \beta_1, t_{ik}, w_{1i}\right) f_Y\left(y_{i\cdot k}|\delta_0, \delta_1, \alpha_W, w_{2i}\right)\right]^{I\left(m_{ik}>0\right)}.$$
$$f_w\left(w_{1i}, w_{2i}, w_{3i}|\Sigma\right) dw_i$$

Because the likelihood function does not have a closed form, either a numerical approximation such as the Gauss-Hermite quadrature[24] or an EM algorithm may be used to estimate the parameters of interest, that is, $(\gamma_0, \gamma_1, \beta_0, \beta_1, \delta_0, \delta_1, \Sigma)$. Gigante[25] developed a numerical estimating procedure for the joint mixed effects model on the number of health insurance payments and the increment payments, in the context of the sleep model, on (M, Y). Specifically, they used a hierarchical likelihood approach[26] and an iterative weighted least square algorithm. However, given the complexity of the joint distribution of (T, M, Y), the use of a numerical algorithm to obtain maximum likelihood parameter estimates can be computationally very intensive and unstable. Thus, we propose to use a stepwise approach that is not only computationally much less intensive and more stable but can also make use of existing software. Initially, Equations (4) to (6) are solved separately as generalized linear mixed effects models to obtain estimates for $(\gamma_0, \gamma_1, \beta_0, \beta_1, \delta_0, \delta_1)$ as well as their variations, the variance parameter for the random effects, that is, $(\sigma_1, \sigma_2, \sigma_3)$. Then we propose three additional steps to estimate the three correlation parameters, that is, $(\rho_{12}, \rho_{13}, \rho_{23})$.

First, we model the number of awakenings without conditioning on the sleep time,

$$\log E\left(M_{ik}|w_i^M\right) = \theta_0^M + \theta_1^M X_{1i} + w_i^M, \tag{8}$$

where it can be shown that $\theta_0^M = \beta_0 + \gamma_0$ and $\theta_1^M = \beta_1 + \gamma_1$; the random effect $w_i^M = w_{1i} + w_{2i}$. Conditional on (w_{1i}, w_{2i}), M_{ik} becomes a negative binomial random variable. Note that without conditioning on the total sleep time, Equation (8) becomes a model on the total number of awakenings instead of on the frequency of awakenings in Equation (5). Both the fixed effects and the random effects on the number of awakenings are additive effects from the effects on the total sleep time and the effects on the rate of awakening. The correlation between (w_{1i}, w_{2i}), ρ_{12}, can then be estimated.

Next, conditioning on (w_{2i}, w_{3i}) and T_{ik}, WASO, $Y_{i\cdot k}$, follows a compound Poisson gamma distribution. We model the mean of WASO using the following mixed effects model[27]

$$\log E\left(Y_{i\cdot k}|\mathrm{T}_{ik}, w_i^Y\right) = \theta_0^Y + \theta_1^Y X_i + \log\left(T_{ik}\right) + w_i^Y, \tag{9}$$

where it can be shown that $\theta_0^Y = \beta_0 + \delta_0$ and $\theta_1^Y = \beta_1 + \delta_1$; the random effect $w_i^Y = w_{2i} + w_{3i}$. Both the fixed effects and the random effects on WASO given the total sleep time are additive effects from the effect on the frequency of awakenings and the effect on the duration of each wake bout. The correlation between (w_{2i}, w_{3i}), ρ_{23}, can then be estimated.

Third, based on Equations (1) to (3), we further model WASO without conditioning on the total sleep time, that is,

$$\log E\left(\mathrm{Y}_{i\cdot k}|w_i^{Y^*}\right) = \theta_0^{Y^*} + \theta_1^{Y^*} X_i + w_i^{Y^*}, \tag{10}$$

where it can be shown that $\theta_0^{Y^*} = \beta_0 + \delta_0 + \gamma_0$ and $\theta_1^{Y^*} = \beta_1 + \delta_1 + \gamma_1$; the random effect $w_i^{Y^*} = w_{1i} + w_{2i} + w_{3i}$. Conditional on (w_{1i}, w_{2i}, w_{3i}), $Y_{i\cdot k}$ follows the compound negative binomial gamma distribution, which is shown to also belong to the Tweedie's family with the index parameter between 1 and 2 (see proof of Lemma in the Appendix). Note that without conditioning on the number of awakenings, Equation (10) becomes a model on WASO instead of a model on each wake bout as Equation (6). Both the fixed effects and the random effects on WASO are additive effects from the effects on the rate of awakenings, on the duration of each wake bout and on the total sleep time. The correlation between (w_{1i}, w_{3i}), ρ_{13}, can be estimated.

Details of the derivation of Equations (8) to (10) and the correlation parameters are provided in the Appendix.

To summarize the estimation procedure, the following steps are needed:

(i) Use Equations (4) to (6) to estimate $(\gamma_0, \gamma_1, \beta_0, \beta_1, \delta_0, \delta_1)$ as well as their variations and $(\sigma_1, \sigma_2, \sigma_3)$.
(ii) Use Equations (8) to (10) to estimate the variation of $(\beta_1 + \gamma_1, \beta_1 + \delta_1, \beta_1 + \delta_1 + \gamma_1)$ and to estimate $(\rho_{12}, \rho_{13}, \rho_{23})$.
(iii) Based on (ii) Cov $\left(\hat{\beta}_1, \hat{\delta}_1\right)$, Cov $\left(\hat{\beta}_1, \hat{\gamma}_1\right)$, and Cov $\left(\hat{\delta}_1, \hat{\gamma}_1\right)$ can be estimated so that joint hypotheses $\beta_1 = \gamma_1 = 0$, $\beta_1 = \delta_1 = 0$, or $\beta_1 = \delta_1 = \gamma_1 = 0$ can be tested via a Wald test.
(iv) Use bootstrap methods to estimate the variability of estimates for Σ.

Equations (4) to (7) can be estimated using an R-package *glmmTMB* and Equations (9) and (10) can be estimated using an R-package *cpglmm*. Program to implement the procedure is available at https://github.com/xiaonanxue/Code/blob/xiaonanxue-sleep-model/Simu%20data%20example%20Xue.R. The performance of this estimation procedure is examined using simulations in Section 3.

2.3 | Sleep efficiency

Another important sleep variable is the sleep efficiency, defined as the proportion of sleep time out of the duration of the rest interval, that is, $Eff_{ik} = {}^{T_{ik}}/_{Y_{i\cdot k}+T_{ik}}$. Because $Y_{i\cdot k} \geq 0$, direct model of sleep efficiency using a mixed beta regression model is not appropriate. A zero-inflated beta-regression is not adequate either because it does not fit our assumption that the number of awakenings follows a regular Poisson model. However, we can use Equation (9) to make inference on sleep efficiency. Specifically, Equation (9) shows that

$$E_Y\left(\frac{1 - Eff_i}{Eff_i}|T_{ik}, w_i^Y\right) = E_Y\left(\frac{Y_{i\cdot k}}{T_{ik}}|T_{ik}, w_i^Y\right) = e^{\theta_0^Y + \theta_1^Y X_i + w_i^Y},$$

where $E_Y\left(\frac{1\text{-}Eff_i}{Eff_i}|T_{ik}, w_i^Y\right)$ is the expected odds of sleep inefficiency, also interpreted as the expected odds of being awake vs being asleep at a night. Thus, $e^{\theta_1^Y}$ represents the ratio in expected odds for sleep inefficiency, or ratio in expected odds for being awake associated with X.

It worth emphasizing that Equations (8) to (10) not only allows the use of existing software to estimate the correlation parameters but each of the questions also provides additional insights on sleep activities. Table 1 summarizes model assumptions and parameter interpretations for each mixed effects model presented in this section.

3 | SIMULATIONS

We used simulation studies to examine the performance of the proposed estimation procedure. For simplicity, we only assume one binary variable X with 50% equal to 1, for example, X is the unemployment status of the participant. First, in each simulated data set, we generated $N = 300$ subjects each observed for 7 nights. This sample size is chosen to be close to our example data set. We considered two scenarios. In the first scenario, $\gamma_1 = 0.0$, $\beta_1 = 0.0$, $\delta_1 = 0.2$, $\sigma_1 = 0.15$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.4$, $\rho_{13} = -0.5$, $\rho_{23} = 0.0$. The fixed effect parameters were chosen to be close to the coefficients for the unemployment status in our data example. The level of random effects were also chosen to be close to our data example (see Table 6). In the second scenario, we set the parameters to be $\gamma_1 = -0.1$, $\beta_1 = 0.2$, $\delta_1 = 0.15$, $\sigma_1 = 0.18$, $\sigma_2 = 0.35$, $\sigma_3 = 0.3$, $\rho_{12} = -0.4$, $\rho_{13} = -0.2$, $\rho_{23} = 0.2$. These parameters are chosen to cover a range of plausible values. The simulation was repeated 500 times and the results are summarized in Table 2.

In the next set of simulations, we examined if the performance of the estimation procedure is influenced by the sample size. We generated $N = 200$ subjects each observed for 7 nights and considered the same two scenarios as above. The results are summarized in Table 3.

In the third set of the simulation, we examined if the performance of the estimation procedure for the joint model is affected by the level of correlations between sleep variables. Therefore, we considered a third scenario with a higher level of correlations, that is, $\gamma_1 = -0.1$, $\beta_1 = 0.2$, $\delta_1 = 0.15$, $\sigma_1 = 0.15$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.5$, $\rho_{13} = -0.6$, $\rho_{23} = 0.4$. We varied the sample size from $N = 300$ to $N = 500$ subjects. The results are summarized in Table 4.

Table 2 shows that the proposed estimating procedure was able to estimate the fixed parameters ($\gamma_1, \beta_1, \delta_1$) accurately with very little bias: the absolute bias <0.001 for $\beta_1 = \gamma_1 = 0$ and the relative bias <1% for δ_1 in scenario 1 and the relative bias <4% for γ_1 and <1% for β_1, δ_1 in scenario 2. The variation of these estimates was estimated accurately with close to 1 in ratio between the averaged estimated SE and the sample SE, also close to nominal coverage for the estimated 95% confidence interval.

Table 2 also shows that the proposed estimation procedure estimates the random effect parameters σ's very well with relative bias <3% and also the correlation parameters ρ's very well: when $\rho_{23} = 0$ in scenario 1, the magnitude of the bias = 0.002; the relative bias <5% in general with one exception: 6.3% for ρ_{12} in scenario 1. Bootstrapping methods were used to estimate the 95% confidence interval for the random effect parameters. Because of the computational intensity, bootstrap samples were set to be 100. Table 2 indicated that the bootstrap confidence intervals are slightly conservative with the coverage probability ranged from 96% to 100%.

When the sample size was reduced to be 200, Table 3 shows that the performance of the procedure estimation procedure was close to the performance when $N = 300$, suggesting that the sample size requirement for the estimation procedure is not extensive. When the correlation between the sleep variables increased, Table 4 shows that the performance of the proposed estimation procedure for the joint model was similar to the performance when the correlations were smaller. We increased the sample size from 300 to 500, both showed good performance, suggesting that the sample size requirement for the proposed estimation procedure is not extensive even when the correlation between sleep variables is high.

4 | APPLICATION

This article was motivated by the WIHS HIV Sleep Study, which enrolled $N = 316$ women aged from 40 to 70 years old in 2018. Participants were asked to wear a wrist actigraphy device continuously for 24 hours a day for 10 days to increase the probability of recording a full 7 days of sleep data. The participants were instructed to press the event marker on the monitor before and after adlib sleep each night and they were also asked to complete a daily written sleep diary.[28]

TABLE 1 Description of mixed effects models on sleep variables

Response	Condition on other metrics	Distribution conditional on random effects	Interpretation of covariate effect on outcome	Model	Fixed effects	Random effects
Total sleep time		Gamma	Ratio in total sleep time	$\log E(T_{ik}\|\upsilon_{1i}) = \gamma_0 + \gamma_1 X_i + w_{1i}$	γ_1	w_{1i}
Number of wake up bouts	Total sleep time	Poisson	Ratio in rate of awakening	$\log E(M_{ik}\|T_{ik}, \upsilon_{2i}) = \beta_0 + \beta_1 X_i + \log(T_{ik}) + w_{2i}$	β_1	w_{2i}
Total wake time	Number of wake up bouts	Gamma	Ratio in wake bout length	$\log E(Y_{i\cdot k}\|M_{ik}, \upsilon_{3i}) = \delta_0 + \delta_1 X_i + \log(M_{ik}) + w_{3i}$	δ_1	w_{3i}
Number of wake up bouts		Negative binomial	Ratio in total number of awakenings	$\log E\left(M_{ik}\|w_i^M\right) = \theta_0^M + \theta_1^M X_i + w_i^M$	$\theta_1^M = \gamma_1 + \beta_1$	$w_i^M = w_{1i} + w_{2i}$
Total wake time	Total sleep time	Compound Poisson gamma	OR of sleep inefficiency	$\log E(Y_{i\cdot k}\|T_{ik}, \upsilon_{1i}, \upsilon_{3i}) = \theta_0^Y + \theta_1^Y X_i + \log(T_{ik}) + w_i^Y$	$\theta_1^Y = \beta_1 + \delta_1$	$w_i^Y = w_{1i} + w_{3i}$
Total wake time		Compound Poisson gamma	Ratio in total wake time	$\log E\left(Y_{i\cdot k}\|w_i^{Y^*}\right) = \theta_0^{Y^*} + \theta_1^{Y^*} X_i + w_i^{Y^*}$	$\theta_1^{Y^*} = \gamma_1 + \beta_1 + \delta_1$	$w_i^{Y^*} = w_{1i} + w_{2i} + w_{3i}$

TABLE 2 Summary results for 500 simulated data each with N subjects and 7 nights of objectively measured sleep variables and a binary exposure variable ($N = 300$, scenarios 1 and 2)

Total sleep time			**Frequency of awakening**			**Length of wake bout**			**Total # of wake bouts**			**Total wake time conditional on total sleep time**			**Total wake time**		
Scenario 1: $\gamma_1 = 0.0$, $\beta_1 = 0.0$, $\delta_1 = 0.2$, $\sigma_1 = 0.15$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.4$, $\rho_{13} = -0.5$, $\rho_{23} = 0.0$																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
Bias[a]	Cov[b] %	$\frac{\text{ave est se}}{\text{sample se}}$[c]	Bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	%[d] bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
0.0003	95.4	0.985	−0.0003	95.4	1.020	0.038	94.4	0.955	−0.014	94.6	0.954	0.321	94.8	0.983	0.602	93.0	0.934
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov[e] %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		% bias	Cov %		Bias[d]	Cov %		% bias	Cov %	
0.972	97.6		0.998	97.2		0.993	98.2		6.342	99.2		−0.002	96.4		0.171	99.2	
Scenario 2: $\gamma_1 = -0.1$, $\beta_1 = 0.2$, $\delta_1 = 0.15$, $\sigma_1 = 0.18$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.4$, $\rho_{13} = -0.2$, $\rho_{23} = 0.2$																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
−3.917	96.4	1.031	−0.636	95.6	1.004	0.434	95.4	1.025	2.340	94.8	0.984	0.338	95.4	1.029	0.739	96.0	1.010
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		**% bias**	Cov %		% bias	Cov %		**% bias**	Cov %	
.986	97.2		0.995	95.0		0.996	97.6		3.989	100		−4.649	98.2		−0.113	99.4	

[a] When parameter = 0, bias = estimated value − true value.
[b] Proportion of 95% confidence intervals that include the true value.
[c] Average of estimated SE of the parameter estimate/sample SE of the parameter estimate.
[d] % bias = (estimated − true value)/true value × 100%.
[e] Bootstrap confidence intervals were used.

TABLE 3 Summary results for 500 simulated data each with N subjects and 7 nights of objectively measured sleep variables and a binary exposure variable ($N = 200$, scenarios 1 and 2)

Total sleep time			**Frequency of awakening**			**Length of wake bout**			**Total # of wake bouts**			**Total wake time conditional on total sleep time**			**Total wake time**		
Scenario 1: $\gamma_1 = 0.0$, $\beta_1 = 0.0$, $\delta_1 = 0.2$, $\sigma_1 = 0.15$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.4$, $\rho_{13} = -0.5$, $\rho_{23} = 0.0$																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
Bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	Bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	Bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
−0.0001	94.8	0.982	0.002	94.8	0.985	0.446	93.8	0.982	−0.002	94.6	0.957	1.105	93.6	0.971	0.671	94.4	0.963
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov[a] %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		% bias	Cov %		Bias	Cov %		% bias	Cov %	
0.962	99.0		0.992	100		0.995	98.8		4.081	99.6		−0.003	96.6		−0.674	98.2	
Scenario 2: $\gamma_1 = -0.1$, $\beta_1 = 0.2$, $\delta_1 = 0.15$, $\sigma_1 = 0.18$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.4$, $\rho_{13} = -0.2$, $\rho_{23} = 0.2$																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
−2.772	93.6	0.973	0.370	95.0	1.002	−1.107	94.4	0.988	0.499	95.2	1.010	−1.528	94.0	0.977	−1.895	93.2	0.957
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov[a] %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		% bias	Cov %		% bias	Cov %		% bias	Cov %	
0.962	99.0		0.993	98.2		0.991	98.6		4.069	97.2		5.310	99.0		−4.390	98.6	

[a] Bootstrap confidence intervals were used.

TABLE 4 Summary results for 500 simulated data each with N subjects and 7 nights of objectively measured sleep variables and a binary exposure variable ($N = 300$ and 500, scenario 3: $\gamma_1 = -0.1$, $\beta_1 = 0.2$, $\delta_1 = 0.15$, $\sigma_1 = 0.15$, $\sigma_2 = 0.35$, $\sigma_3 = 0.30$, $\rho_{12} = -0.5$, $\rho_{13} = -0.6$, $\rho_{23} = 0.4$)

Total sleep time			**Frequency of awakening**			**Length of wake bout**			**Total # of wake bouts**			**Total wake time conditional on total sleep time**			**Total wake time**		
***N* = 300 subjects**																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
−1.190	95.0	1.012	0.040	95.6	1.035	−1.448	95.2	1.011	1.254	96.6	1.071	−0.571	95.6	1.027	−0.680	96.6	1.035
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov[a] %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		% bias	Cov %		% bias	Cov %		% bias	Cov %	
0.982	99.0		0.995	93.8		0.995	98.8		4.196	98.6		−3.298	99.0		−1.780	96.0	
***N* = 500 subjects**																	
Fixed effects parameters																	
γ_1			β_1			δ_1			$\beta_1 + \gamma_1$			$\beta_1 + \delta_1$			$\beta_1 + \delta_1 + \gamma_1$		
% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$	% bias	Cov %	$\frac{\text{ave est se}}{\text{sample se}}$
−1.519	94.6	1.001	0.315	94.2	1.001	0.300	93.6	0.974	0.963	95.2	1.005	−0.854	94.2	0.971	−1.748	95.4	1.005
Random effects parameter																	
σ_1			σ_2			σ_3			ρ_{12}			ρ_{23}			ρ_{13}		
$\bar{\hat{\sigma}}_1/\sigma_1$	Cov[a] %		$\bar{\hat{\sigma}}_2/\sigma_2$	Cov %		$\bar{\hat{\sigma}}_3/\sigma_3$	Cov %		% bias	Cov %		% bias	Cov %		% bias	Cov %	
0.985	98.8		1.004	94.0		1.000	99.0		4.209	98.2		3.064	99.2		−1.625	98.4	

[a] Bootstrap confidence intervals were used.

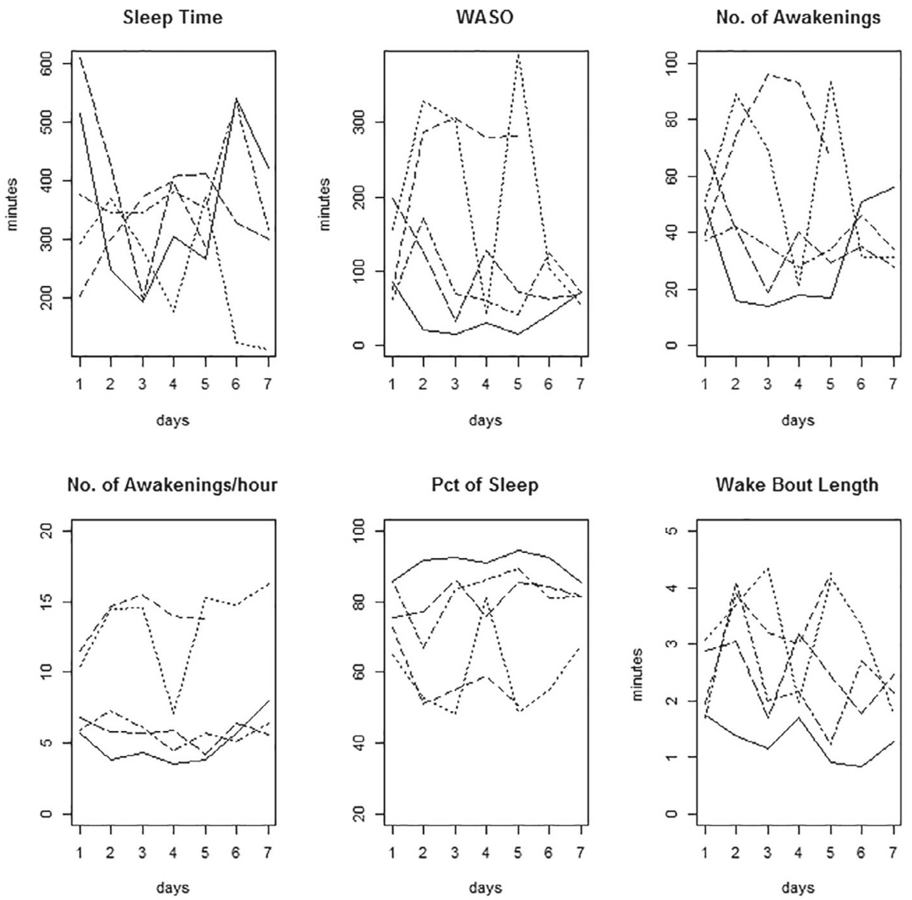

FIGURE 1 Total sleep time, WASO, total number of awakenings, number of awakenings per hour of sleep (ie, awake frequency), sleep efficiency (percent of sleep time), and mean wake bout length over multiple days for five randomly selected participants from the WIHS HIV Sleep Study

Actigraphy recordings were analyzed with the Actiware 6.0.9 program (Respironics, Bend, OR). The setting of nightly rest intervals was guided by event markers, sleep diaries, light data and activity level.[10] The actigraphy readings were interpreted up to the first 7 days by Dr. Burgess's research team (Sleep and Circadian Research Laboratory, University of Michigan). We limited the study sample to 291 women who had at least 3 consecutive days of sleep data. Among them, the majority of the women (91%) had 7 days of actigraphy data. Objective measured sleep variables were then derived from the first up to 7 days of available actigraphy recordings. Figure 1 displays total sleep time, WASO, total number of awakenings, number of awakenings per hour of sleep (ie, awake frequency), sleep efficiency (percent of sleep) and mean bout length over up to 7 days for five participants randomly selected from the study population. Figure 1 shows that there was a large variability within a participant over multiple nights and the variability across participants appeared to be larger for wake patterns than that for total sleep time.

Summary statistics of the sleep variables averaged over multiple days per subject are provided in Table 5. Figure 2 shows pairwise scatterplots and Spearman correlations between individual sleep variables averaged over multiple days for each participant. As indicated in Figure 2, the averaged sleep time was slightly positively associated with the average number of awakenings; and it was negatively associated with the average frequency of awakenings and the average length of wake bouts, their Spearman correlations are −0.37 and −0.22, respectively. As expected, average WASO was positively correlated with the average number of awakenings and the average frequency of awakenings as well as the average length of wake bouts. The average number of awakenings was positively correlated with the average frequency of awakenings but was not associated with the average wake bout lengths. The average frequency of awakenings appeared to be positively correlated with the average wake bout lengths with a small correlation of 0.16.

In this article, we are interested in examining how the participants' unemployment status (unemployed = 1, part time or full time working = 0) and age influence their total sleep time, WASO, frequency and number of awakenings and

TABLE 5 Summary statistics of sleep variables for the WIHS HIV Sleep Study with $N = 291$ women each with 3 to 7 nights of Actigraphy measured sleep variables and averages over nights are used for each participant

	Mean	Median	First quartile	Third quartile	(Min, Max)
Total sleep time (in minutes)	382.30	379.60	336.60	427.70	(183.10, 591.20)
Number of awakenings	35.66	34.00	27.71	41.57	(7.10, 81.86)
Number of awakenings per hour sleep	5.76	5.46	4.40	6.69	(1.60, 15.38)
Total wake time (in minutes)	66.29	62.00	42.57	81.21	(13.14, 226.92)
Sleep efficiency (%)	85.63	86.32	82.57	90.09	(59.80, 96.75)
Average wake bout length (in minutes)	1.85	1.69	1.37	2.15	(0.81, 5.02)

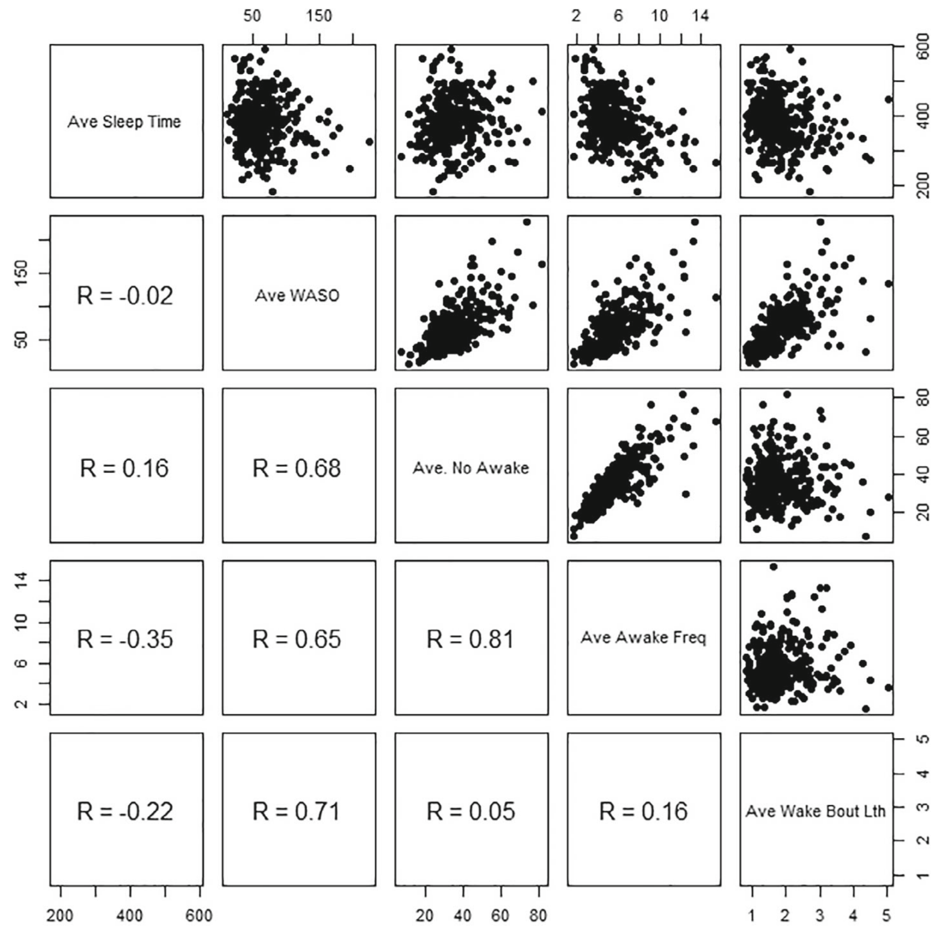

FIGURE 2 Pairwise scatter plot and Spearman correlations between total sleep time, total wake time, total number of awakenings, number of awakenings per hour of sleep (ie, awake frequency), and mean bout length over averaged over multiple days per participant in the WIHS HIV Sleep Study

wake bout lengths and how these sleep variables are related to each other. About 64% of women was unemployed. We dichotomized age at ≥50 or <50 years: 65% of women were over 50 years of age.

4.1 | Parameter estimates

The proposed method was applied to the study population and the results are summarized in Table 6. Table 6 shows that compared to part time or full time employed participants, unemployed participants did not have less sleep time, neither did they wake up more frequently but they tended to have a longer wake bout (ratio in bout length = 1.161 [95% CI: 1.070, 1.259; $P < .001$]). Consequently, the unemployed participants had a longer WASO (ratio in length = 1.184 [95% CI: 1.055, 1.331; $P = .004$]), which resulted from a longer duration in each wake episode. Unemployed status was associated with sleep inefficiency with an odds ratio (OR) of 1.176 (95% CI: 1.031, 1.340; $P = .016$), that is, unemployed participants were more likely to stay awake during their rest interval. Women in the older age group (≥50 years old) did not sleep less nor did they wake up more frequently, but they also tended to have a longer wake bout (ratio in bout length = 1.090 [95% CI: 1.002, 1.186; $P = .044$]). As a result, older age was also associated with a longer WASO (ratio in length = 1.135 [95% CI: 1.007, 1.280; $P = .038$]). However, older age was not associated with sleep inefficiency, in another words, older aged participants were not more likely to stay awake during their rest interval. Wald test on $\beta_1 = \delta_1 = 0$ and $\beta_1 = \delta_1 = \gamma_1 = 0$ for unemployment was significant, demonstrating an overall effect of unemployment on sleep activities in various dimensions. Older age only had a borderline joint effect on sleep variables (P-value = .090 for $\beta_2 = \delta_2 = \gamma_2 = 0$). We also examined HIV status, BMI and waist circumference, none of which demonstrated any association with any of these sleep variables.

Not much heterogeneity in total sleep time was observed among the participants with the SD for the random effect estimated to be 0.147 (95% CI: 0.112, 0.165). This implies that a subject whose underlying level of sleep duration is at 75th percentile of the population sleeps 20% longer than a subject whose underlying level is at 25th percentile, given the same characteristics otherwise. However, substantial heterogeneity was observed in number of awakenings and wake bout durations: the estimated SD for the random effect associated with the number of awakenings conditional on total sleep time and with the duration of each wake bout is 0.353 (95% CI: 0.319, 0.384), 0.320 (95% CI: 0.292, 0.346), respectively. For example, the frequency of awakenings for a subject whose underlying tendency to wake up is at 75th percentile of the population is approximately 60% higher than that for a subject whose underlying level is at 25th percentile, given that they have the same characteristics otherwise. Similar level of heterogeneity was observed on duration of wake bouts. In summary, we found that heterogeneity in these sleep variables among our study participants primarily lies in wakening patterns, which confirmed our observation in Figure 1.

Random effects associated with the total sleep time were negatively correlated with random effects for the frequency of awakenings: $\rho_{12} = -0.406(95\%\text{CI} : -0.551, 0.242)$ and also were negatively correlated with random effects for wake bout durations $\rho_{13} = -0.534(95\%\text{CI} : -0.720, -0.366)$. Random effects for the frequency of awakenings were not correlated with random effects for wake bout durations: $\rho_{23} = 0.011(95\%\text{CI} : -0.153, 0.144)$. These findings were in general consistent with what we observed in Figure 2, which suggest that someone who tends to wake-up more frequently or tends to stay awake longer each wake episode tends to sleep less, having an overall poorer sleep quality.

4.2 | Model fitting

To assess how our model fits to the data, we examined the standardized Pearson residuals for models (4) to (6) to detect any large discrepancy between observed and fitted daily sleep variables. As shown in Figure 3, for models (4) and (6), there were very few residuals whose magnitude exceeded 2.5; for model (5), there were 3.4% of residuals exceeded 2.5, suggesting no significant discrepancy between observed and fitted sleep variables. We further compared the empirical distribution between each subject's observed sleep variables and its corresponding fitted value, both averaged over multiple nights for each participant. Figure 3 shows that the empirical distributions for the averaged observed and fitted number of awakenings are very close and the distributions for the averaged observed and fitted WASO are almost identical. Figure 3 also shows that the distribution of averaged fitted values for the total sleep time is less spread than the distribution of the observed value, suggesting that there is extra variability that the current model is not able to explain and additional risk factors may be needed to explain the high values of the total sleep time in particular.

5 | DISCUSSION

To our knowledge, for the first time, we proposed to jointly model several sleep variables including total sleep time, number of awakenings and WASO. A joint model allows separate as well as overall assessment of risk factor influences on several sleep variables as well as evaluation of relationships among sleep variables. An estimating procedure was developed based upon estimating a series of mixed effects models, which avoids computational intensity and instability that may arise in numerical search of maximum likelihood estimates. The estimation procedure can be implemented

TABLE 6 Joint analysis of total sleep time, number of awakenings, WASO, and sleep time of the WIHS HIV Sleep Study with $N = 291$ women with 3 to 7 nights of objectively measured sleep variables

	Ratio in total sleep time e^{γ}(CI)	Ratio in frequency of awakening e^{β}(CI)	Ratio in length of wake bout e^{δ}(CI)	Ratio in total # of wake bouts $e^{\beta+\gamma}$(CI)	Ratio in odds of sleep inefficiency $e^{\beta+\delta}$(CI)	Ratio in WASO $e^{\beta+\delta+\gamma}$(CI)
Unemployed	1.011 (0.963, 1.061) $P = 0.645$	1.017 (0.928, 1.115) $P = 0.714$	1.161 (1.070, 1.259) $P < .001$	1.027 (0.944, 1.116) $P = .542$	1.176 (1.031, 1.340) $P = .016$	1.184 (1.055, 1.331) $P = .004$
	$\beta_1 = \gamma_1 = 0, P = .746$[a]		$\beta_1 = \delta_1 = 0, P = .002$		$\beta_1 = \delta_1 = \gamma_1 = 0, P = .001$	
Age >50	1.020 (0.970, 1.072) $P = .442$	1.015 (0.923, 1.116) $P = .757$	1.090 (1.002, 1.186) $P = .044$	1.036 (0.950, 1.130) $P = .422$	1.101 (0.962, 1.261) $P = .162$	1.135 (1.007, 1.280) $P = .038$
	$\beta_2 = \gamma_2 = 0, P = .584$		$\beta_2 = \delta_2 = 0, P = .132$		$\beta_2 = \delta_2 = \gamma_2 = 0, P = .090$	
	σ_1 Est (CI)	σ_2 Est (CI)	σ_3 Est (CI)	$\sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2$ Est (CI)	$\sigma_2^2 + \sigma_3^2 + 2\rho_{23}\sigma_2\sigma_3$ Est (CI)	$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\rho_{12}\sigma_1\sigma_2 + 2\rho_{13}\sigma_1\sigma_3 + 2\rho_{23}\sigma_2\sigma_3$ Est (CI)
Random effects	0.147 (0.112, 0.165)[b]	0.353 (0.319, 0.384)	0.320 (0.292, 0.346)	0.323 (0.287, 0.357)	0.479 (0.422, 0.521)	0.401 (0.347, 0.442)
	ρ_{12}		ρ_{13}		ρ_{23}	
	−0.406 (−0.551, −0.242)		−0.534 (−0.720, −0.366)		0.011 (−0.153, 0.144)	

[a] Based on Wald test.
[b] Bootstrap confidence interval.

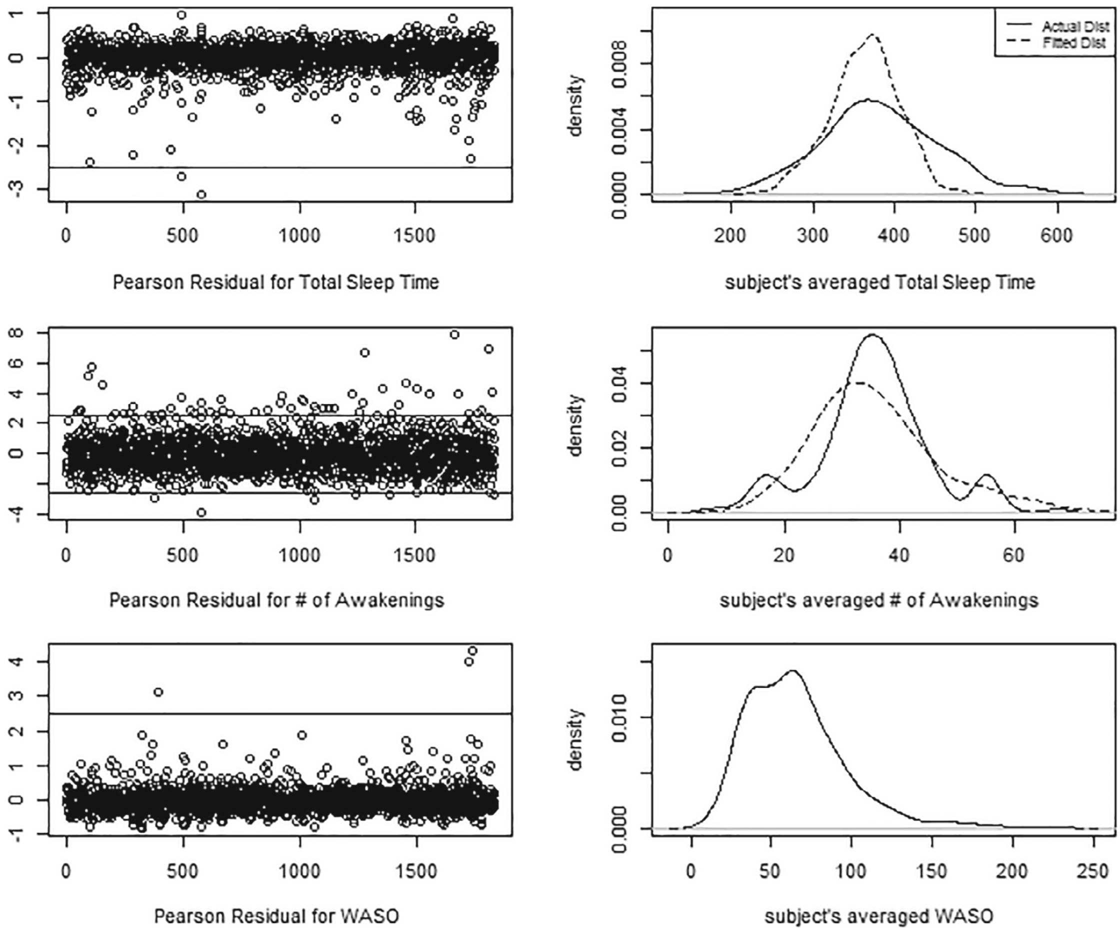

FIGURE 3 Model fitting for number of wake-up bouts (top), wake time (middle), and total sleep time (bottom). Left panel is the standardized Pearson residuals and the right panel is the comparison between the empirical distribution for each sleep metric averaged over nights under observation and the empirical distribution for its corresponding fitted value

using existing packages in R. Furthermore, each mixed effects model provides a unique interpretation for the effects of risk factors on sleep activities.

We applied the proposed methods to an ongoing HIV Sleep Study to examine factors that may affect participants' sleep. We found that women who were unemployed or older had a longer WASO, resulted from a longer duration of wake bouts rather than more frequent awakenings at night. Neither of these two variables affected the total sleep duration. Only unemployment status but not older age resulted in decreased sleep efficiency. A large amount of heterogeneity was observed in frequency and duration of wake bouts. Total sleep time was negatively associated with frequency and duration of awakenings. Thus, someone who woke up more frequently or someone who stayed awake longer each wake episode tended to sleep less. Findings from this study help better understand the inter-relationship between sleep variables and can help the development of interventions in order to improve sleep quality overall.

Our models fit the data generally well except that the fitted distribution of total sleep time appeared to have less variability than the observed distribution. Large heterogeneity remained unexplained in the number of awakenings and WASO. These results suggest that additional variables may be needed to better understand these sleep variables. For example, in addition to sociodemographic variables, daytime napping or environmental factors such as light or noise may also influence night sleep activities. Furthermore, while sleep onset is random, sleep end may not be because the participant may not wake up naturally. Thus, some of the total sleep time can be right censored. This makes it more challenging to model the total sleep time.

Several limitations of our proposed method warrant further study. First, our model assumes that conditional on personal heterogeneity, a sleep variable measured repeatedly over multiple nights is independent from each other. However, it is possible that a subject's sleep activities in one night affect his/her sleep activities during the next night. Thus, serial correlations may need to be incorporated using a transition model. In addition, there is also growing evidence that the stability and timing of the sleep-wake cycle are important predictors of cardio-metabolic risk.[15] Models to include the night-to-night variability in sleep variables and the timing of sleep onset, sleep ends and sleep mid-point as well as the timing of wake time (ie, early or late during the rest interval) will provide a more comprehensive assessment of overall sleep quality.

ACKNOWLEDGEMENTS

The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH). MWCCS (Principal Investigators): Atlanta CRS (Ighovwerha Ofotokun, Anandi Sheth, and Gina Wingood), U01-HL146241; Baltimore CRS (Todd Brown and Joseph Margolick), U01-HL146201; Bronx CRS (Kathryn Anastos and Anjali Sharma), U01-HL146204; Brooklyn CRS (Deborah Gustafson and Tracey Wilson), U01-HL146202; Data Analysis and Coordination Center (Gypsyamber D'Souza, Stephen Gange and Elizabeth Golub), U01-HL146193; Chicago-Cook County CRS (Mardge Cohen and Audrey French), U01-HL146245; Chicago-Northwestern CRS (Steven Wolinsky), U01-HL146240; Northern California CRS (Bradley Aouizerat, Jennifer Price, and Phyllis Tien), U01-HL146242; Los Angeles CRS (Roger Detels and Matthew Mimiaga), U01-HL146333; Metropolitan Washington CRS (Seble Kassaye and Daniel Merenstein), U01-HL146205; Miami CRS (Maria Alcaide, Margaret Fischl, and Deborah Jones), U01-HL146203; Pittsburgh CRS (Jeremy Martinson and Charles Rinaldo), U01-HL146208; UAB-MS CRS (Mirjam-Colette Kempf, Jodie Dionne-Odom, and Deborah Konkle-Parker), U01-HL146192; UNC CRS (Adaora Adimora), U01-HL146194. The MWCCS is funded primarily by the National Heart, Lung, and Blood Institute (NHLBI), with additional co-funding from the *Eunice Kennedy Shriver* National Institute of Child Health & Human Development (NICHD), National Institute on Aging (NIA), National Institute of Dental & Craniofacial Research (NIDCR), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Neurological Disorders and Stroke (NINDS), National Institute of Mental Health (NIMH), National Institute on Drug Abuse (NIDA), National Institute of Nursing Research (NINR), National Cancer Institute (NCI), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute on Minority Health and Health Disparities (NIMHD), and in coordination and alignment with the research priorities of the National Institutes of Health, Office of AIDS Research (OAR). MWCCS data collection is also supported by UL1-TR000004 (UCSF CTSA), UL1-TR003098 (JHU ICTR), UL1-TR001881 (UCLA CTSI), P30-AI-050409 (Atlanta CFAR), P30-AI-073961 (Miami CFAR), P30-AI-050410 (UNC CFAR), P30-AI-027767 (UAB CFAR), and P30-MH-116867 (Miami CHARM).

The authors gratefully acknowledge the contributions of the study participants and dedication of the staff at the MWCCS sites.

This work is supported by National Heart, Lung, and Blood Institute (NHLBI) 1R01HL142116-01 (2018-2022) entitled "The Indoleamine 2, 3-dioxygenase (IDOZe) Study" (Audrey, Burgess).

The authors would also want to thank the two anonymous referees whose constructive suggestions have resulted in significant improvement of the article.

DATA AVAILABILITY STATEMENT

The data used in this article is not publicly available because it is part of an ongoing research study, as research data are not shared.

ORCID

Xiaonan Xue https://orcid.org/0000-0003-2872-8119

REFERENCES

1. Patel SR, Malhotra A, White DP, Gottlieb DJ, Hu FB. Association between reduced sleep and weight gain in women. *Am J Epidemiol*. 2006;164:947-954. doi:10.1093/aje/kwj280
2. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. *Lancet*. 1999;354:1435-1439. doi:10.1016/S0140-6736(99)01376-8
3. Yaggi HK, Araujo AB, McKinlay JB. Sleep duration as a risk factor for the development of type 2 diabetes. *Diabetes Care*. 2006;29:657-661. doi:10.2337/diacare.29.03.06.dc05-0879
4. Cespedes EM, Dudley KA, Sotres-Alvarez D, et al. Joint associations of insomnia and sleep duration with prevalent diabetes: the Hispanic community health study/study of Latinos (HCHS/SOL). *J Diabetes*. 2016;8:387-397. doi:10.1111/1753-0407.12308

5. Aurora RN, Kim JS, Crainiceanu C, et al. Habitual sleep duration and all-cause mortality in a general community sample. *Sleep*. 2016;39:1903-1909. doi:10.5665/sleep.6212
6. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. *Sleep*. 2006;29:1155-1173.
7. de Souza L, Benedito-Silva AA, Pires MLN, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. *Sleep*. 2003;26:81-85.
8. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;2:389-396.
9. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003;26:342-392.
10. Patel SR, Weng J, Rueschman M, et al. Reproducibility of a standardized actigraphy scoring algorithm for sleep in a U. S. Hispanic/Latino population. *Sleep*. 2015;38:1497-1503. doi:10.5665/sleep.4998
11. Bathgate CJ, Edinger JD, Wyatt JK, Krystal AD. Objective but not subjective short sleep duration associated with increased risk for hypertension in individuals with insomnia. *Sleep*. 2016;39:1037-1045. doi:10.5665/sleep.5748
12. Ramos AR, Weng J, Wallace DM, et al. Sleep patterns and hypertension using actigraphy in the Hispanic community health study/study of Latinos. *Chest*. 2018;153:87-93. doi:10.1016/j.chest.2017.09.028
13. Franck LS, Johnson LM, Lee K, et al. Sleep disturbances in children with human immunodeficiency virus infection. *Pediatrics*. 1999;104:e62. doi:10.1542/peds.104.5.e62
14. Bacon MC, von Wyl V, Alden C, et al. The Women's interagency HIV study: an observational cohort brings clinical sciences to the bench. *Clin Diagn Lab Immunol*. 2005;12:1013-1019. doi:10.1128/CDLI.12.9.1013-1019.2005
15. Abbott SM, Weng J, Reid KJ, et al. Sleep timing, stability, and BP in the Sueño ancillary study of the Hispanic community health study/study of Latinos. *Chest*. 2019;155:60-68. doi:10.1016/j.chest.2018.09.018
16. Luik AI, Zuurbier LA, Hofman A, et al. Stability and fragmentation of the activity rhythm across the sleep-wake cycle: the importance of age, lifestyle, and mental health. *Chronobiol Int*. 2013;30:1223-1230. doi:10.3109/07420528.2013.813528
17. Chen J, Patel SR, Redline S, et al. Weekly sleep trajectories and their associations with obesity and hypertension in the Hispanic/Latino population. *Sleep*. 2018;41:zsy150. doi:10.1093/sleep/zsy150
18. Mihaylova B, Briggs A, O'Hagan A, et al. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011;20:897-916. doi:10.1002/hec.1653
19. Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*. 2005;6:93-109. doi:10.1093/biostatistics/kxh020
20. Smyth GK, Jørgensen B. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *ASTIN Bull*. 2002;32:143-157. doi:10.2143/AST.32.1.1020
21. Kurz CF. Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Med Res Methodol*. 2017;17:171. doi:10.1186/s12874-017-0445-y
22. Jørgensen B. Exponential dispersion models. *J R Stat Soc B Methodol*. 1987;49:127-145. doi:10.1111/j.2517-6161.1987.tb01685.x
23. Delong Ł, Lindholm M, Wüthrich MV. Making Tweedie's compound Poisson model more accessible. *Eur Actuar J*. 2021;11:185-226. doi:10.1007/s13385-021-00264-3
24. Liu Q, Pierce DA. A note on Gauss-Hermite quadrature. *Biometrika*. 1994;81:624-629. doi:10.2307/2337136
25. Gigante P, Picech L, Sigalotti L. A mixture model for payments and payment numbers in claims reserving. *ASTIN Bull*. 2016;48:25-53.
26. Lee Y, Nelder JA. Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*. 2001;88:987-1006.
27. Zhang Y. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Stat Comput*. 2013;23:743-757. doi:10.1007/s11222-012-9343-7
28. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*. 2012;35:287-302. doi:10.5665/sleep.1642

How to cite this article: Xue X, Hua S, Weber K, et al. Jointly modeling of sleep variables that are objectively measured by wrist actigraphy. *Statistics in Medicine*. 2022;41(15):2804-2821. doi: 10.1002/sim.9385

APPENDIX . DERIVATION OF EQUATIONS (8) TO (10) AND THE CORRELATION PARAMETERS

Let $T_{ik}^S = \frac{T_{ik}}{\alpha_S v_{1i} r_i} \mid v_{1i} \sim \text{Gamma}\,(\alpha_S, 1/\alpha_S)$, then based on Equation (4),

$$\log T_{ik} = \gamma_0 + \gamma_1 X_i + \log\left(T_{ik}^S\right) + w_{1i},$$

where $T_{ik}^{S} \perp w_{1i}$. Based on Equations (1) and (2),

$$E\left(M_{ik}|v_{1i}, v_{2i}\right) = E_T\left(E\left(M_{ik}|T_{ik}, v_{1i}, v_{2i}\right)\right) = E_T\left(\lambda_i v_{1i} v_{2i} T_{ik}\right).$$

Thus, with Equation (5), we have

$$\log E\left(M_{ik}|w_{1i}, w_{2i}, T_{ik}^{S}\right) = \beta_0 + \gamma_0 + (\beta_1 + \gamma_1) X_{1i} + \log\left(T_{ik}^{S}\right) + w_{1i} + w_{2i}.$$

Because $T_{ik}^{S} \perp v_{1i}$ and $T_{ik}^{S} \perp v_{2i}$, conditional on (v_{1i}, v_{2i}), M_{ik} is a Poisson random variable with over-dispersion such that $\mathrm{Var}\left(M_{ik}|v_{1i}, v_{3i}\right) = E\left(M_{ik}|v_{1i}, v_{3i}\right)\left(1 + 1/\alpha_S E\left(M_{ik}|v_{1i}, v_{3i}\right)\right)$. We therefore model the number of awakenings without conditioning on the sleep time,

$$\log E\left(M_{ik}|w_i^M\right) = \theta_0^M + \theta_1^M X_{1i} + w_i^M, \tag{8}$$

which becomes a mixed effect negative binomial model. We have $\theta_0^M = \beta_0 + \gamma_0$ and $\theta_1^M = \beta_1 + \gamma_1$; the random effect $w_i^M = w_{1i} + w_{2i}$. The correlation between (w_{1i}, w_{2i}), $\rho_{12} = \frac{\mathrm{var}(w_i^M) - \sigma_1^2 - \sigma_2^2}{2\sigma_1\sigma_2}$ can then be estimated.

Next, conditioning on (w_{2i}, w_{3i}), $Y_{ijk} \perp M_{ik}$, $Y_{i \cdot k} \sim$ compound Poisson gamma. As we mentioned earlier, the compound Poisson gamma distribution belongs to the Tweedie's family and does not have a closed form of density function. We model the mean of WASO using the following mixed effects model[27]

$$\log E\left(Y_{i \cdot k}|\mathrm{T}_{ik}, w_i^Y\right) = \theta_0^Y + \theta_1^Y X_i + \log\left(T_{ik}\right) + w_i^Y. \tag{9}$$

As $E_Y\left(Y_{i \cdot k}|T_{ik}, v_{2i}, v_{3i}\right) = E_M\left(E_Y\left(Y_{i \cdot k}|T_{ik}, M_{ik}, v_{2i}, v_{3i}\right)\right) = \lambda_i \alpha_W \xi_i T_{ik} v_{2i} v_{3i}$, we have $\theta_0^Y = \beta_0 + \delta_0$ and $\theta_1^Y = \beta_1 + \delta_1$; the random effect $w_i^Y = w_{2i} + w_{3i}$. The correlation between (w_{2i}, w_{3i}), $\rho_{23} = \frac{\mathrm{var}(w_i^Y) - \sigma_2^2 - \sigma_3^2}{2\sigma_2\sigma_3}$ can then be estimated.

Third, based on Equations (1) to (3),

$$E_Y\left(Y_{i \cdot k}|v_{1i}, v_{2i}, v_{3i}\right) = E_T E_Y\left(Y_{i \cdot k}|T_{ik}, v_{1i}, v_{2i}, v_{3i}\right) = E_T\left(\lambda_i \alpha_W \xi_i T_{ik} v_{2i} v_{3i}|v_{1i}, v_{2i}, v_{3i}\right) = \lambda_i \alpha_W \alpha_S \xi_i r_i v_{1i} v_{2i} v_{3i}.$$

We further model WASO without conditioning on the total sleep time, that is,

$$\log E\left(\mathrm{Y}_{i \cdot k}|w_i^{Y^*}\right) = \theta_0^{Y^*} + \theta_1^{Y^*} X_i + w_i^{Y^*}. \tag{10}$$

We showed in below lemma that the compound negative binomial gamma distribution also belongs to the Tweedie's family with the index parameter between 1 and 2. We then have $\theta_0^{Y^*} = \beta_0 + \delta_0 + \gamma_0$ and $\theta_1^{Y^*} = \beta_1 + \delta_1 + \gamma_1$; the random effect $w_i^{Y^*} = w_{1i} + w_{2i} + w_{3i}$. Since $\mathrm{var} w_i^{Y^*} = \mathrm{var} w_i^M + \mathrm{var} w_i^Y - \mathrm{var} w_{2i} + 2\rho_{13}\sigma_1\sigma_3$, the correlation between (w_{1i}, w_{3i}), $\rho_{13} = \frac{\mathrm{var}\left(w_i^{Y^*}\right) - \mathrm{var}\left(w_i^Y\right) - \mathrm{var}\left(w_i^M\right) + \sigma_2^2}{2\sigma_1\sigma_3}$ can be estimated.

Lemma. *Assume $M \mid T \sim \mathrm{Poisson}(\lambda T)$, $T \sim \mathrm{Gamma}\left(\alpha_S, 1/\alpha_S\right)$, $Y_j \sim \mathrm{Gamma}(\alpha, \gamma)$, Y_j's are iid and $M \perp Y_j$ s.t. $Y = \sum_{j=1}^{M} Y_j$ follows a Tweedie's distribution.*

Proof. Marginally, M follows a negative binomial distribution so that $E(M) = \lambda$, and $\mathrm{Var} M = \lambda\left(1 + 1/\alpha_S \lambda\right)$. Therefore, $EY = \mu = \lambda\alpha\gamma$, $\mathrm{Var} Y = \mu\alpha\gamma\left(\frac{1}{\alpha} + 1 + \frac{\lambda}{\alpha_S}\right)$. Then

$\mathrm{Var} Y = \varphi\mu^p$ where $p = 1 + \frac{\alpha_S + \lambda\alpha}{(\alpha+1)\alpha_S + \lambda\alpha}$, and $\varphi = \frac{\lambda^{1-p}(\alpha\gamma)^{2-p}}{2-p}$.

It can be easily seen that $p \in (1, 2)$. ∎