Submitted to the Annals of Applied Statistics

BAYESIAN ESTIMATION OF PHYSICAL ACTIVITY FROM WEARABLE SENSOR MEASUREMENTS

BY JIWON LEE[1,a], YOUNGDEOK HWANG[2,d] WON CHANG[1,b], BO YOUNG HONG[3,e], AND HANG J. KIM[1,c]

[1]*Division of Statistics and Data Science, University of Cincinnati,* [a]*lee9j2@mail.uc.edu;* [b]*changwn@ucmail.uc.edu;* [c]*kim3h4@ucmail.uc.edu*

[2]*Paul H. Chook Department of Information Systems and Statistics Baruch College, City University of New York,* [d]*youngdeok.hwang@baruch.cuny.edu*

[3]*Department of Rehabilitation Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea,* [e]*byhong@catholic.ac.kr*

It has become a common practice to use various wearable devices not only to monitor personal health but also to conduct clinical studies. We introduce a new Bayesian approach to analyze the data collected from personal devices in which measurements are repeatedly taken from the wearer. The proposed approach is designed to process data for individual wearers and combine them for further statistical inference, such as between-group analysis. It is based on modeling to filter the measurements of the active period from noise measurements of the resting periods so that the activity characteristics are properly captured. To accommodate a time-dependent nature with the switching periods, the proposed model utilizes a temporal kernel structure along with reversible jump Markov chain Monte Carlo. The approach is illustrated with a case study based on three-axis accelerometer activity monitors. An `R` package for the proposed methodology is provided in an open repository.

1. Introduction. The use of accelerometer-based wearable devices for physical activity recognition has gained popularity. These devices use pre-defined algorithms to process raw signal data to generate outcomes for further analysis. Accelerometers measure changes in speed and direction of body movements continuously over time, which allows an in-depth analysis of physical activities with various metrics, such as total movement, step counts, and intensity of physical activity (Conradsson et al., 2017; Summers et al., 2019; Leavy et al., 2021).

However, analyzing such a dataset poses statistical challenges. One major issue is that algorithms processing the outcomes often rely on heuristic designs that require thorough validation. For example, a commonly investigated metric from wearable devices is the number of steps recorded, where a threshold needs to be determined for acceleration so that a measurement can be counted as a step. This conversion is incorporated by comparing data collected in advance under controlled laboratory environments, where exact step counts can be obtained (John et al., 2018). The criteria used to classify vigorous physical activity from resting state is based on cut-off values from previous studies, but the values vary across studies (Kim, Beets and Welk, 2012). These algorithms hardly apply to cases involving a diverse population with physical impairments, as we considered in our case study. The algorithms used in commercial devices to convert accelerometer measurements to steps, distance, and energy are often proprietary and not shared with researchers (Wright et al., 2017).

Another challenge faced by practitioners is the lack of available analysis methodology. While there are tools developed to process and analyze raw datasets from various devices

Keywords and phrases: Actigraphy, Automated sensors, Reversible-jump Markov chain Monte Carlo.

(Choi et al., 2021), these are primarily designed for validation studies (Choi et al., 2011) and may not be suitable for a thorough statistical analysis. This absence becomes particularly challenging when analyzing the outcomes associated with additional information from wearers, or pre-obtained data in the laboratory is not available or applicable. For example, the participants are often asked to complete a diary in sleep studies (van Hees et al., 2018), yet discrepancies have been reported between the sleep diary and the device (Lawrence and Muza, 2018). Physical activity studies often require participants to visit a facility in advance so that a training data can be obtained to calibrate the activity classification (Trost, Zheng and Wong, 2014; Bianchim et al., 2023). This limitation diminishes the advantage of wearable devices, where non-intrusive and continuous data collection are spotlighted as the benefits.

We introduce a novel Bayesian approach that circumvents these challenges by analyzing the data from wearable devices with the minimal need for additional information. This approach is designed to estimate the latent states of the subject, identifying whether a recorded measurement is from an active or resting state. Once the measurements are labeled with their estimated states, practitioners can use them for further analysis of physical activities. In addition to in-depth analyses, the proposed method provides uncertainty quantification associated with the estimated latent states as the proposed model is constructed for fully probabilistic inference. Clinical researchers can make a better-informed decision with the degree of uncertainty. The proposed approach also furnishes rigorous statistical inference for parameter estimates within the Bayesian framework.

There are three major contributions in our work. First, our modeling approach calibrates the labeling estimation by considering the individual subject's varying behavioral characteristics. This is particularly useful when the subjects are from a diverse population that is different from the majority, i.e., high-functioning adult population, as demonstrated in our case study of pediatric subjects with mobility challenges. Second, we tailored a reversible jump Markov chain Monte Carlo (RJMCMC, Green 1995) approach to solve a latent state inference problem through a change-point detection method. Unlike the original form of the RJMCMC, change point estimation in our problem requires not only latent state labeling but also retaining the alternating structure between active and resting states. Third, we propose an intuitive framework for a clinical study to compare different groups after subject-level data have been individually processed by our proposed algorithm. This framework facilitates a large-scale comparison study by aggregating the computational results from smaller subsets.

The remainder of the article is organized as follows. Section 2 introduces the data used for the case study and reviews the relevant literature. Section 3 introduces our model and describes the estimation procedure. Section 4 illustrates the proposed method with an application to a clinical study using an accelerometer. We conclude with a short summary and discussion in Section 5. Posterior derivation is given in the Supplementary Material, and an `R` package, `BayesActigraph`, is provided in an open repository for practitioners to implement the proposed methodology.

2. Background. Studies using actigraph devices examine people's behavior by analyzing the length of sedentary time, sleeping pattern, and active time, continuously recorded for an extended period. We use the data collected from a wearable device called *Actigraph wGT3X-BT*, a compact $(4.6 \times 3.3 \times 1.5$ cm) and lightweight $(19$ g) tri-axial accelerometry-based activity monitor that can be worn on the wrist, waist, ankle, or thigh. Acceleration measurements are taken in three-dimensional directions and recorded every minute in standard gravity units (G, 9.8 m/s^2). These measurements are then converted to the Euclidean norm of a three-dimensional vector, known as the vector magnitude (VM).

Table 1 presents a sample dataset from a study participant in our application to a clinical study. Measurements were collected over approximately a week for each subject, with slight variations, resulting in approximately 10,000 records for each subject.

TABLE 1
A snapshot of the sample dataset from a study participant

Time Stamp	axis1	axis2	axis3	VM
2017-09-06 15:30:00	116	0	56	129
2017-09-06 15:31:00	126	32	218	254
2017-09-06 15:32:00	845	663	1213	1621

The dataset was originally obtained for a medical study of pediatric subjects with cerebral palsy. A total of 58 subjects with varying levels of movement ability were observed, where the most severely impaired case includes the group that cannot independently walk and use powered mobility. We used VM as our primary measurement for model development and analysis. Each subject has additional demographic information, including age, gender, and movement ability measured in the Gross Motor Function Classification System, a widely used clinical metric for individuals with cerebral palsy (Palisano et al., 1997). There were three groups of subjects: ambulatory, marginally ambulatory, and non-ambulatory, where 26, 16, and 16 subjects belonged to each group, respectively. The proportion, overall average, and variability of non-zero measurements vary by group, reflecting the characteristics of the subject groups with physical disabilities. Figure 1 exhibits snapshots from two subjects in a clinical study involving cerebral palsy. Subject 2's mobility is more severely impacted, which is reflected in the lower level of VM activity compared to Subject 1. The figure shows their distinct patterns in physical mobility, but is insufficient to provide quantitative evaluations and an individualized calibration to take into account varying levels of physical mobility.

Despite its potential for various applications with granular data, statistical analysis for actigraphy data often remains limited to comparing descriptive summary statistics (e.g., Yoon et al., 2022) and is constrained by limited data accessibility to practitioners and a lack of suitable statistical tools. Our specific aim is to devise a methodology addressing two major challenges. First, we incorporate the discrepancy between study data and manufacturer-provided calibration thresholds. Device calibration typically relies on healthy and fully functioning adults, while clinical studies often examine a diverse population, including those who are physically impaired. Hence, it is imperative to develop a model adaptable to individual physical characteristics. Second, the model should facilitate statistical inference for various metrics beyond simple descriptive statistics. Practitioners seek a more comprehensive understanding of participants' physical activities, involving comparisons of various characteristics, for example, frequency, duration, and intensity of physical activity across different groups. Our modeling framework is designed to provide a thorough analysis, addressing the aforementioned objectives.

3. Model and Computation. This section describes the Bayesian model used for state labeling and its inference procedures tailored to wearable sensor measurements. We assume that there are two states alternating: active and resting. In the active state, dynamic physical movements are expected, such as walking and running, while the resting state comprises sleep and resting periods with little body movement. These two states are not observed; they are the latent (hidden) states.

The hidden Markov model (Rabiner, 1989), which is a popular model in the hidden state inference problem, is not applicable to our problem because its strong Markovian assumption is not suitable for wearable devices measurements which often have long-range dependence. The conditional random fields model, an extension of hidden Markov models with less strict Markovian assumption using the features extracted from data (Lafferty, McCallum and Pereira, 2001), is not applicable either, as the dataset has limited features to use. Moreover, perhaps most importantly, these models should be trained in a supervised manner, and hence require additional sources of information such as sleep diaries.

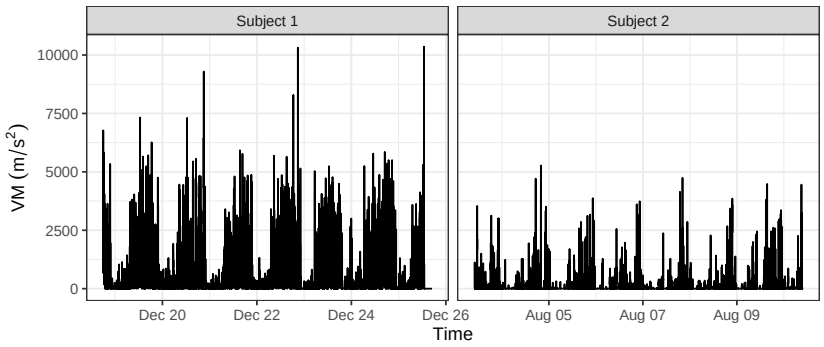

FIG 1. *Illustrative examples of the data from two subjects, where Subject 1's data shows higher activities than Subject 2, and its active and resting states are more visually discernible than those of Subject 2.*

To address these challenges, we propose an interval-based state labeling inspired by the change-point detection methodology of Green (1995). A pair of points, the start and end points, constitutes an interval labeled as either an active or a resting state. Multiple change points mark the transitions between the two states. Each measurement between these points is then assigned the same state label, depending on the interval to which it belongs.

3.1. *Model Specification.* We start the model specification with defining the set of change points denoted by $\mathcal{S} = \{a_1, r_1, \ldots, a_B, r_B\}$ with the specific ordering $r_{b-1} < a_b < r_b < a_{b+1}$, where $A_b = (a_b, r_b]$ and $R_b = (r_b, a_{b+1}]$ represent the b-th active and resting intervals, respectively. Because of the constraints, $r_{b-1} < a_b < r_b < a_{b+1}$, in defining the change points, the B active intervals and the $B-1$ resting intervals always alternate, i.e., no active or resting intervals are adjacent to a same state interval. Let y_t denote the observed value at time $t \in \mathcal{T}^y$ where $\mathcal{T}^y = \{1, \ldots, T\}$. The total of T measurements $\{y_t : t \in \mathcal{T}^y\}$ are assigned active or resting intervals. Let $\boldsymbol{y}_a = \{y_t : t \in A_b,\ b = 1, \ldots, B\}$ denote the collection of measurements in active states and $\boldsymbol{y}_r = \{y_t : t \in R_b,\ b = 1, \ldots, B-1\}$ the collection of measurements in resting states.

3.1.1. *Model for Active State Measurements.* We assume that the measurements during the active states $\boldsymbol{y}_a$ follow a multivariate normal distribution $\boldsymbol{y}_a \sim N(\mu\mathbf{1}, \sigma^2\boldsymbol{\Sigma})$. To incorporate the temporal dependence between measurements, we consider the (absolute) exponential kernel for $\boldsymbol{\Sigma}$ whose (t, t')-th element is $\exp(-|t - t'|/l) + g\delta_{t,t'}$ where $\delta_{t,t'} = 1$ for $t = t'$ and zero otherwise. The kernel matrix $\boldsymbol{\Sigma}$ depends solely on the time difference and the range parameter l (> 0) with the nugget effect g fixed as a small number called jitter (Neal, 1999). This setting mitigates the computational burden associated with the kernel matrix, as only two parameters σ^2 and l need to be estimated. We use a constant mean model, but it can be extended to include other covariates $\boldsymbol{x}^T$ if available. The temporal kernel component ensures that a few abrupt drops in the measurements do not immediately trigger a change in the state from active to resting. Its presence makes the transition process less sensitive than the manufacturer-setting threshold for the measurement level, ensuring the state labeling is robust.

3.1.2. *Model for Resting State Measurements.* The measurements recorded during the resting states $\boldsymbol{y}_r$ have the following characteristics: (a) having excessive zeros, (b) being heavily skewed to the right, and (c) having sporadic spikes. This nature of $\boldsymbol{y}_r$, especially with the presence of spikes, makes it difficult to be modeled using standard distributions.

A Poisson distribution can be considered, but a single Poisson distribution is not flexible enough to handle spikes occurring at varying scales. We instead use a mixture of Poisson distributions for $\boldsymbol{y}_r$ with K mixture components, which gives the likelihood function for $y_t \in \boldsymbol{y}_r$ as $\sum_{k=1}^{K} \pi_k \text{Pois}(y_t; \lambda_k)$, where $\pi_k = p(z_t = k)$ for $k = 1, \ldots, K$, z_t represents a membership indicator of the mixture components, and $\text{Pois}(y_t; \lambda_k)$ is the Poisson likelihood evaluated at y_t with parameter λ_k.

3.1.3. *Prior Distributions.* We use a flat prior for μ and uniform priors with wide ranges for the kernel parameters l and σ^2 involved in the likelihood for $\boldsymbol{y}_a$ to reflect vague prior knowledge about them. For $\boldsymbol{y}_r$, we assume a Dirichlet process prior for a mixing measure G and adopt its stick-breaking representation, i.e., G has a general form of $G = \sum_{k=1}^{K} \pi_k \delta_{\lambda_k}$ (Ishwaran and James, 2001). Note that G governs $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$. In the stick-breaking representation, $\boldsymbol{\pi}$ is expressed as $\pi_k = U_k \prod_{j=1}^{k-1}(1 - U_j)$ for $k = 2, \ldots, K-1$ and $\pi_1 = U_1$ where $U_j \sim \text{Beta}(1, \alpha)$ for $j = 1, \ldots, K-1$ and $U_K = 1$. The Poisson parameters λ_k's are assumed to follow a base distribution G_0. We specify G_0 as a Gamma distribution with its shape parameter α_0 and rate parameter β_0 to achieve the conjugate relationship with the Poisson distributions. This model is known as a Dirichlet process mixture model with a truncated stick-breaking construction. Our model is truncated at K, yet a Dirichlet process theoretically allows infinite-dimensional generalization (Teh, 2010). In practice, by choosing a sufficiently large K, we can ensure that the number of non-empty mixture components stays well below K and hence the approximation error is minimized.

3.2. *Posterior Inference Computation.* Let $\Theta = \left\{\mu, \sigma^2, l\right\}$ and $\Phi = \{\boldsymbol{\lambda}, \boldsymbol{\pi}, \mathcal{Z}\}$ denote a set of parameters for $\boldsymbol{y}_a$ and $\boldsymbol{y}_r$, respectively, where $\mathcal{Z}$ is a set of membership indicators z_t. We have three parameter sets: Θ, Φ, and $\mathcal{S}$ with the joint posterior distribution $f(\Theta, \Phi, \mathcal{S} | \boldsymbol{y}_a, \boldsymbol{y}_r)$. Estimating $f(\Theta | \mathcal{S}, \boldsymbol{y}_a)$ and $f(\Phi | \mathcal{S}, \boldsymbol{y}_r)$ can be implemented with standard MCMC methods with Metropolis-Hastings. On the other hand, in estimating $\pi(\mathcal{S} | \Theta, \Phi, \boldsymbol{y}_a, \boldsymbol{y}_r)$, not only the locations but also the number of the change points in $\mathcal{S}$ can change, i.e., the dimension of $\mathcal{S}$ is not fixed. To address this challenge, we adopt RJMCMC (Green 1995). The key aspect of RJMCMC is that it allows a Markov chain to traverse parameter spaces of different dimensions. This is accomplished within the Metropolis-Hastings framework by allowing a proposal state of a Markov chain that changes the dimension of $\mathcal{S}$. Algorithm 1 summarizes the MCMC implementation. When initializing the state of $\mathcal{S}$, all measurements less than a small cut-off value are assigned to resting states. This threshold

Algorithm 1 Algorithm for MCMC implementation

Initialize $\boldsymbol{\Theta}^{(1)} = \left\{\mu, \sigma^2, l\right\}$, $\boldsymbol{\Phi}^{(1)} = \{\boldsymbol{\lambda}, \boldsymbol{\pi}, \mathcal{Z}\}$, and $\mathcal{S}^{(1)} = \{a_1, r_1, \ldots, a_B, r_B\}$
for m in $2, \ldots, M$ **do**
  Update Θ **and** Φ
    $\Theta^{(m)} \sim \pi\left(\Theta^{(m-1)} | \boldsymbol{y}_a\right)$ ▷ Supplement 1.1
    $\Phi^{(m)}$ is obtained via the blocked Gibbs sampling ▷ Supplement 1.2
  Update $\mathcal{S}$
    Sample $u \sim \mathcal{U}(0, 1)$
    Update $\mathcal{S}^{(m)}$ via
      if $u \leq p_{update}$ **then** update move
      else if $u \leq p_{update} + p_{birth}$ **then** birth move
      else death move
      end if
end for

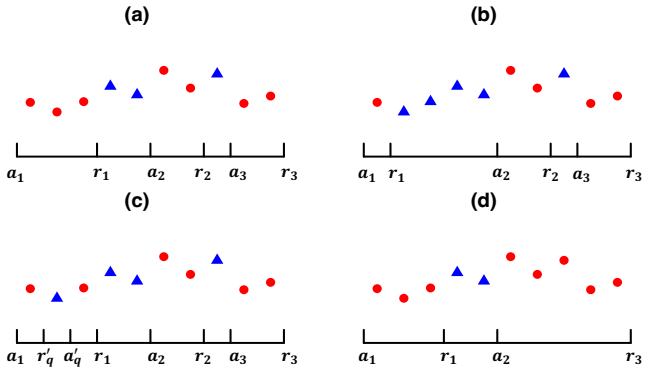

FIG 2. *Illustration of update, birth, and death moves when* $T = 10$ *and* $B = 3$*: (a) current state of* $\mathcal{S}$*, (b) update move for* r_1*, resulting in the change of the state of* y_2 *and* y_3 *to the resting, (c) birth move, creating a new resting interval* $(r'_q, a'_q]$*, leading to the change in the state of* y_2*, and (d) death move for the resting interval* $(r_2, a_3]$*, causing* y_8 *to transition to the active state.*

aligns with the consensus that lower-valued measurements are more likely to be in the resting states. At each iteration, Θ is updated by the Metropolis-Hastings, and Φ is updated by the blocked Gibbs sampling (Ishwaran and James, 2001). We describe the details of the MCMC steps for updating Θ and Φ in Section S1 of the supplementary material.

Possible moves with respect to $\mathcal{S}$ are: (1) changing the location of change points a_b and r_b (*update*), (2) creating a new active or resting interval (*birth*), and (3) removing an existing active or resting interval (*death*). Figure 2 illustrates each move. Note that the birth move increases the number of change points in the current state $\mathcal{S}$ by 2, whereas the death move decreases it by 2. At each iteration, we proceed with one of the three moves – update, birth, or death moves – with probabilities p_{update}, p_{birth}, and p_{death}, respectively. Note that after the birth or death move is accepted, appropriate label changes are needed in the remaining elements of $\mathcal{S}$. For example, if the death move of q-th active interval $(a_q, r_q]$ is accepted, a_b and r_b are relabeled to a'_{b-1} and r'_{b-1} for $b = q+1, \ldots, B$.

We describe each move type in greater detail below. Here $f_A(\cdot)$ and $f_R(\cdot)$ denote the likelihood functions for $\boldsymbol{y}_a$ and $\boldsymbol{y}_r$, respectively, as described in Section 3.1. For simplicity, we assume that state transitions occur at the midpoint between two observation time stamps. This assumption leads to the support of $\mathcal{S}$ to become $\mathcal{T}^S = \{t - 0.5 : 1 \le t \le T + 1\}$.

3.2.1. *Update Move.* The update move changes the locations of $r_1, a_2, \ldots, r_{B-1}, a_B \in \mathcal{S}$, while fixing the locations of the first and last points for simplicity, i.e., a_1 and r_B are fixed as 0.5 and $T + 0.5$, respectively, throughout the entire MCMC iterations. Thus, we only consider updating $2B - 2$ points among $2B$ change points in $\mathcal{S}$.

The prior distribution of $\mathcal{S}$ is given by $\pi(\mathcal{S}) = (2B-2)!/(T-1)^{(2B-2)}$ which is the probability mass function of $2B - 2$ order statistics $(r_1, \ldots, a_B)$ from a discrete uniform distribution $U[1.5, T - 0.5]$. The setting of $\pi(\mathcal{S})$ assigns the same prior probability to the locations of the change points over the support while forcing them to satisfy the ordered constraint of $r_{b-1} < a_b < r_b < a_{b+1}$. For each b, we update a_b (or r_b) as follows: (i) propose a_b^* (or r_b^*) randomly chosen from the proposal distribution $g(a_b^*) \sim U[r_{b-1} + 1, r_b - 1]$ (or $g(r_b^*) \sim U[a_b + 1, a_{b+1} - 1]$), (ii) accept a_b^* (or r_b^*) with the acceptance probability, $\alpha_u = \min\{1, \gamma_u\}$. The probability $\gamma_u = [f_A(\boldsymbol{y}^*_{a'}) f_R(\boldsymbol{y}^*_{r'})] \,/\, [f_A(\boldsymbol{y}_{a'}) f_R(\boldsymbol{y}_{r'})]$ where $\boldsymbol{y}^*_{r'} = \{y_t : r_{b-1} < t \le a_b^*\}$, $\boldsymbol{y}^*_{a'} = \{y_t : a_b^* < t \le r_b\}$, and $\boldsymbol{y}_{r'} = \{y_t : r_{b-1} < t \le a_b\}$ when considering a_b^*; $\boldsymbol{y}^*_{a'} = \{y_t : a_b < t \le r_b^*\}$, $\boldsymbol{y}^*_{r'} = \{y_t : r_b^* < t \le a_{b+1}\}$, and $\boldsymbol{y}_{r'} = \{y_t : r_b < t \le a_{b+1}\}$

when considering r_b^*; and $\boldsymbol{y}_{a'} = \{y_t : a_b < t \leq r_b\}$ with the same form when considering both a_b^* and r_b^*. Note that this proposal scheme is designed to satisfy the constraint $r_{b-1} < a_b < r_b < a_{b+1}$, but the proposal probability is independent of a_b (or r_b) and only dependent of the two endpoints of the uniform distribution, $\{r_{b-1}+1, r_b-1\}$ when considering a_b^* (or $\{a_b+1, a_{b+1}-1\}$ when considering r_b^*). This results in the proposal ratio $g(a_b)/g(a_b^*)$ (or $g(r_b)/g(r_b^*)$) being equal to one. Moreover, the prior $\pi(\mathcal{S})$ only depends on B and T, which remain unchanged by the proposal $\mathcal{S}'$. Thus, the acceptance probability α_u only consists of the likelihood ratio as in γ_u.

3.2.2. *Birth Move.* With the current state $\mathcal{S} = \{a_1, r_1, \ldots, a_B, r_B\}$, the birth move creates a new interval whether an active interval $(a'_q, r'_q]$ or a resting interval $(r'_q, a'_{q+1}]$. To preserve the alternating structure of the two states, each interval still satisfies the constraint $r_{b-1} < a_b < r_b < a_{b+1}$ with the newly added interval. Here, we only introduce the birth of $(a'_q, r'_q]$, and the birth of $(r'_q, a'_{q+1}]$ can be constructed in a similar way.

Let $(r_{q-1}, a_q]$ be a randomly chosen resting interval in which the addition of a new active interval $(a'_q, r'_q]$ is suggested such that $r_{q-1} < a'_q < r'_q < a_q$. Then, the proposed state of $\mathcal{S}$ is defined by $\mathcal{S}' = \left\{a'_1, r'_1, \cdots, a'_{B+1}, r'_{B+1}\right\}$ where $a'_b = a_b$ and $r'_b = r_b$ for $b = 1, \ldots, q-1$; and $a'_{b+1} = a_b$ and $r'_{b+1} = r_b$ for $b = q, \ldots, B$. The acceptance probability for $\mathcal{S}'$ is $\alpha_b = \min\{1, \gamma_b\}$ where $\gamma_b = \left[f_A(\boldsymbol{y}_q)\pi(\mathcal{S}')p\left(\mathcal{S}' \to \mathcal{S}\right)\right] / \left[f_R(\boldsymbol{y}_q)\pi(\mathcal{S})p\left(\mathcal{S} \to \mathcal{S}'\right)\right] |J|$, decomposed into three parts: (1) posterior ratio $\left[f_A(\boldsymbol{y}_q)\pi(\mathcal{S}')\right] / \left[f_R(\boldsymbol{y}_q)\pi(\mathcal{S})\right]$, (2) proposal ratio $p\left(\mathcal{S}' \to \mathcal{S}\right))/p\left(\mathcal{S} \to \mathcal{S}'\right)$, and (3) Jacobian $|J|$. See Green (1995); Green and Mira (2001); Fan and Sisson (2011) for the complete derivation. When $(a'_q, r'_q]$ is created, the likelihood changes only for the observations $\boldsymbol{y}_q = \{y_t : a'_q < t \leq r'_q\}$ which are evaluated with f_R under $\mathcal{S}$ but with f_A under $\mathcal{S}'$. This constructs the posterior ratio as $\left[f_A(\boldsymbol{y}_q)\pi(\mathcal{S}')\right] / \left[f_R(\boldsymbol{y}_q)\pi(\mathcal{S})\right]$ with the prior $\pi(\mathcal{S})$ defined in the previous update move and $\pi(\mathcal{S}') = (2B)!/(T-1)^{2B}$.

Let $p\left(\mathcal{S}' \to \mathcal{S}\right)/p\left(\mathcal{S} \to \mathcal{S}'\right)$ denote the proposal ratio where $p\left(\mathcal{S} \to \mathcal{S}'\right)$ is a probability of proposing $\mathcal{S}'$ from $\mathcal{S}$ and $p\left(\mathcal{S}' \to \mathcal{S}\right)$ is its reverse. The latter will be discussed in the next part as it involves a death move of removing $(a'_q, r'_q]$. The probability $p\left(\mathcal{S} \to \mathcal{S}'\right)$ consists of three parts involved in proposing $\mathcal{S}'$. First, the birth move is chosen with p_{birth}. Second, we select the interval $(r_{q-1}, a_q]$ where the birth is attempted with the probability proportional to its length, i.e., $\pi_q^R = L_q^R / \sum_{b=2}^{B} L_b^R$ where L_b^R denote the length of $(r_{b-1}, a_b]$. The rationale behind this is that longer intervals are more likely to contain mislabeled y_t. Thus, breaking down a long interval into smaller pieces is preferred for efficiency. Third, we need to match the dimensions of $\mathcal{S}$ and $\mathcal{S}'$ to make the Markov chain reversible. One way to achieve this is to generate two random variables from a known joint distribution and link them to a'_q and r'_q using a deterministic function h such that $h(u_1, u_2) = (a'_q, r'_q)$. We use the identity function h, so that $h(u_1, u_2) = (u_1, u_2) = (a'_q, r'_q)$. Since $r_{q-1} < a'_q < r'_q < a_q$, we assume that (u_1, u_2) are two order statistics of random variables independently drawn from $U[r_{q-1}+1, a_q-1]$, that is $p(a'_q, r'_q) = p(u_1, u_2) = 2!(a_q - r_{q-1} - 1)^{-2}$. Following the three probabilities involved, $p\left(\mathcal{S} \to \mathcal{S}'\right) = p_{birth} \times \pi_q^R \times 2!(a_q - r_{q-1} - 1)^{-2}$. The Jacobian term arises due to the transformation of (u_1, u_2) to (a'_q, r'_q), and it is one as we use the identity function.

Then, the acceptance probability for the birth of $(a'_q, r'_q]$ is $\alpha_b = \min\{1, \gamma_b\}$ where $\gamma_b = \left[f_A(\boldsymbol{y}_q)/f_R(\boldsymbol{y}_q)\right]\left[(2B)(2B-1)(T-1)^{-2}\right]\left[p\left(\mathcal{S}' \to \mathcal{S}\right)/p_{birth}\, \pi_q^R\, 2!(a_q - r_{q-1} - 1)^{-2}\right]$. We defer the explanation of $p\left(\mathcal{S}' \to \mathcal{S}\right)$ to the next paragraph, which is the probability of proposing the "death" move.

3.2.3. *Death Move.* The death move combines three intervals to one, e.g., choosing an active interval $(a_b, r_b]$ then combining it with its neighboring resting intervals $(r_{b-1}, a_b]$ and

TABLE 2
Summary of the five largest estimated mixing weights $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ and their corresponding Poisson parameters λ_k.

k		1	2	3	4	5
$\boldsymbol{\pi}$	Posterior mean	0.666	0.063	0.051	0.049	0.048
	95% CI	(0.652, 0.681)	(0.056, 0.069)	(0.045, 0.058)	(0.043, 0.055)	(0.043, 0.055)
$\boldsymbol{\lambda}$	Posterior mean	0.005	21.0	40.69	68.04	7.37
	95% CI	(0.003, 0.008)	(20.3, 21.8)	(39.5, 41.8)	(66.6, 69.6)	(7.0, 7.9)

$(r_b, a_{b+1}]$, resulting to a new resting interval $(r'_{b-1}, a'_b]$ where $r'_{b-1} = r_{b-1}$ and $a'_b = a_{b+1}$. When we consider the death of an active interval $(a_q, r_q]$, the proposed state is $\mathcal{S}' = \{a'_1, r'_1, \cdots, a'_B, r'_B\}$ where $a'_b = a_b$ and $r'_b = r_b$ for $b = 1, \ldots, q-1$; and $a'_b = a_{b+1}$ and $r'_b = r_{b+1}$ for $b = q, \ldots, B-1$. We chose an interval for the death move with probability inversely proportional to its lengths, i.e., $\pi_q^A = L_q^A / \sum_{b=2}^{B-1} L_b^A$ where L_b^A denote the length of $(a_b, r_b]$. Unlike the birth move, this ensures that shorter intervals are assigned higher probabilities of selection for removal, i.e., a short active interval is merged by neighboring resting intervals or vice versa. Given that data from wearable devices are typically collected at high frequency, such as every 30 seconds or 1 minute, an immediate switch from one state to another can be considered redundant partitioning. The probability of proposing the death of an interval is written by $p(\mathcal{S} \to \mathcal{S}') = p_{death} \times \pi_q^A$.

When the active interval $(a_q, r_q]$ is removed, the likelihood changes for the observations $\boldsymbol{y}_q = \{y_t : a_q < t \leq r_q\}$ which are evaluated with f_A under $\mathcal{S}$ but with f_R under $\mathcal{S}'$. Then, the acceptance probability for the death move is $\alpha_d = \min\{1, \gamma_b\}$ where $\gamma_d = \left[f_R(\boldsymbol{y}_q)\pi(\mathcal{S}')p_{birth}\ \pi_q^R\ 2!(a_{q+1} - r_{q-1} - 1)^{-2}\right] / \left[f_A(\boldsymbol{y}_q)\pi(\mathcal{S})\ p_{death}\ \pi_q^A\right]$. For the death moves, the relevant probabilities are defined by $\pi_q^R = (a_{q+1} - r_{q-1}) / \sum_{b \neq q-1, q}^{B} L_b^R$ and $\pi(S') = (2B-4)!/(T-1)^{2B-4}$.

4. Case Study. We apply our model to the dataset from a study involving pediatric subjects with cerebral palsy as introduced in Section 2. We first present the result from an individual subject, demonstrating how the latent labeling algorithm performs on real data. Some additional detail for computation is also provided. We then illustrate group-level comparison analysis, using additional metric analysis derived from individual subject's posterior. Some remarks on formulating general group-level comparisons are also provided.

4.1. *Application on an Individual Subject.* For each subject, we run MCMC described in Algorithm 1 for $M = 8,000$ iterations after a burn-in period of 2,000 iterations, with the maximum number of mixture components K set as 25 (see Section 23.3 of Gelman et al., 2013, for setting the number of mixture components in a DP mixture model). We set G_0 for the resting state to follow a Gamma distribution, $G_0 \sim \text{Gamma}(\alpha_0, \beta_0)$, where its prior mean α_0/β_0 equals the first quantile of all non-zero measurements from the group to which a subject belongs. Choosing the first quantile is based on the intuition that the measurements in the resting state tend to be low. We set the prior variance α_0/β_0^2 ten times greater than the prior mean to accommodate individual heterogeneity by setting $\beta_0 = 0.1$. When such group information is not available, one can either use individual information or prior domain knowledge.

Table 2 further summarizes the posterior distribution of $\boldsymbol{\pi}$ and corresponding $\boldsymbol{\lambda}$ used for the mixture of Poisson distributions modeling the resting state. Close-to-zero observations in the resting state can be explained by the first mixture component with an estimated posterior

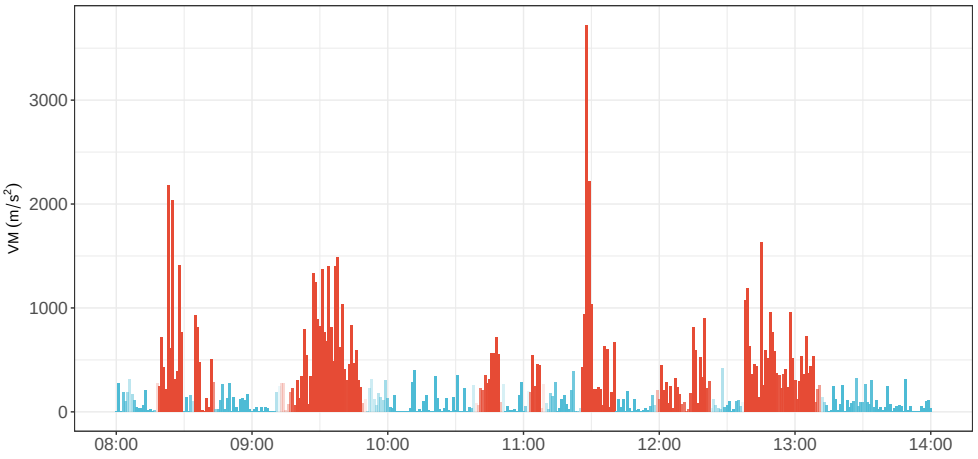

FIG 3. *An example of the result of the state-labeling applied to the dataset of a 6-hour period from a subject with x-axis being the time index and y axis the VMs. Red color indicates the active states and blue the resting. Color opacity represents label uncertainty.*

mean of 0.005, which accounts for (on average over the posterior draws) about 66.6% of measurements labeled as the resting state. The remaining components spread out to capture occasional spikes possibly observed in the resting state, preventing them from being labeled as active states. To incorporate the state uncertainty, we calculate the posterior active state probability of y_t, $\hat{\rho}_t = M^{-1}\sum_{m=1}^{M}\sum_{b=1}^{\hat{B}_m}\mathbb{1}\left(y_t \in \hat{A}_{mb}\right)$ for $t = 1, \ldots, T$, where $\hat{B}_m$ and $\hat{A}_{mb}$ represents the estimated number of active intervals and the b-th active interval at mth MCMC iteration, respectively, as defined in Section 3.1.

Figure 3 displays the result of the state-labeling for a subject with the VMs recorded over a six-hour period. It depicts the result with y_t in red if $\hat{\rho}_t \geq 0.5$, and blue otherwise, with the state uncertainty indicated by color opacity, i.e., darker red indicates y_t whose $\hat{\rho}_t$ is closer to one. Two main observations can be made here. First, this posterior probability $\hat{\rho}_t$ reflects the uncertainty of the state-labeling, where the lighter colors are particularly observed near the boundaries, indicating increased uncertainties associated with state transitions. Unlike the proprietary algorithms to process the same data, our model can provide a rigorous and transparent uncertainty measure for its result. Second, it visually demonstrates that both active and resting states are robust to noisy observations of the VMs. In the resting state, the VMs are primarily zero, yet some isolated instances of high VMs are observed, similar to those observed in the middle of an active period. These incidents, however, are still labeled as resting period and treated as one-off measurement noise. This behavior differs from the threshold-based approach that uses a single cut-off value to divide the resting and active states.

Figure 4 depicts the average active minutes for a given time period to examine a potential diurnal pattern. We first divide all the measurements divided into 24 one-hour intervals, and calculate the proportion of the measurements that are labeled as active states. Once the proportions for each interval are calculated, credible intervals can be easily calculated using MCMC samples. The black line in Figure 4 connects the posterior mean in minutes for each hour interval, where the upper and lower bounds of 95% credible interval are represented in shades. The narrow credible bands for the period corresponding to early morning indicate the consistent sleep-wake pattern.

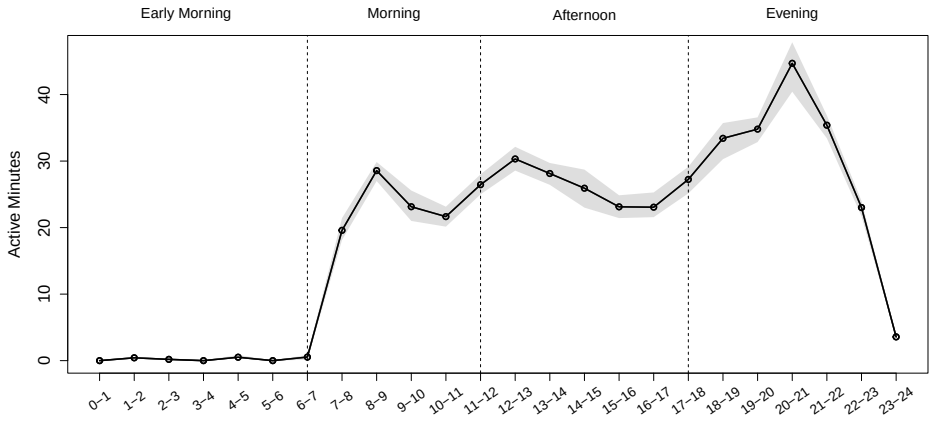

FIG 4. *Visualization of the daily activity pattern, where the x-axis and y-axis represent the 24 time periods and average active minutes, respectively. The black solid line connects the posterior means with the shaded area covering the 95% credible intervals at a given time.*

4.2. *Group Level Comparison Analysis.* The discussion so far has focused on the inference on the individual subject, where subject-level inference is obtained without any coordination of data with other subjects. Now we shift toward the subsequent group-level inference aggregating the subject-level posterior samples. The group-level inference primarily concerns aggregating the individual RJMCMC posterior samples, instead of utilizing a fully joint model. This simplified approach allows a convenient overview of individuals at the aggregate level in a broader perspective. We outline a procedure for making a group-level comparison with two metrics: active endurance and daily total duration of intense activity. The overall idea can be easily extended to more general use cases. For simplicity, we describe each metric for a single subject.

First, we examine the active endurance. The active endurance measures how long one can sustain activity on average, expressed in minutes. To give its definition, let $\hat{S}_m = \left\{\hat{a}_{m1}, \hat{r}_{m1}, \ldots, \hat{a}_{m\hat{B}_m}, \hat{r}_{m\hat{B}_m}\right\}$ denote the estimated change points at m-th iteration. For $m = 1, \ldots, M$, we calculate the m-th posterior sample of the active endurance as $E_m = \sum_{b=1}^{\hat{B}_m}(\hat{r}_{mb} - \hat{a}_{mb})/\hat{B}_m$. Since individuals with limited mobility usually cannot sustain physical activity for an extended time, examining this metric is expected to shed light on studying the endurance of the subjects with different levels of cerebral palsy.

The second aspect we examined is the daily total duration of intense activity. Using the model in Section 3.1, we define a VM measurement as intense when it exceeds its mean μ. We calculate the proportion of all the measurements that exceed μ, which is then converted to a daily average in hours. This conversion accounts for the varying data availability for each subject.

Two proposed measures can be used as alternatives to Moderate-to-vigorous physical activity (MVPA) that is a common measure for describing the intensity of activities, which categorizes a subject's activity into three categories: moderate, vigorous, and very vigorous (Nader et al., 2008). A limitation of conventional MVPA is that it uses predefined cut-off values for all subjects indiscriminately, which is not adequate for the study involving many subjects with mobility challenges like our case study. Instead, our intense activity naturally adjusts to each subject's personal mobility level.

Each point in Figure 5 depicts the posterior means of active endurance and intense activities for each subject, with the horizontal and vertical bars corresponding to their 95% credible intervals. Different symbols correspond to three groups of different mobility, where Group

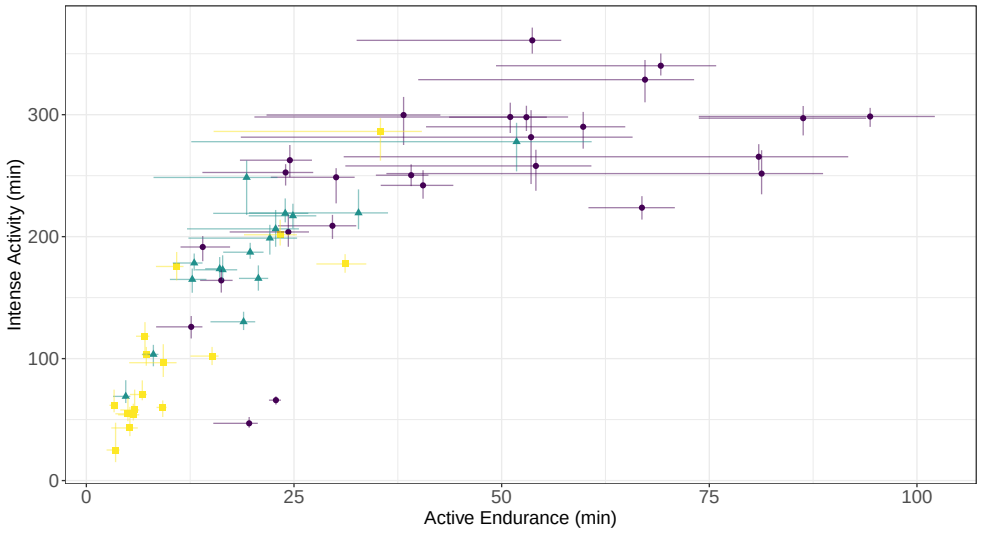

FIG 5. *Posterior means and 95% credible intervals of active endurance and intense activity for each subject in minutes, where Group 1, Group 2, and Group 3 are represented by circle, triangle, and square symbols, respectively.*

1 represents those with the mildest mobility challenges and Group 3 represents the most severe. On average, physical competence in both intense endurance and activity duration aligns well with the level of mobility. Subjects in Group 1 tend to have higher variation in active duration during MCMC iterations, reflected by longer horizontal bars than other two groups. This observation suggests that many measurements' labels are switching between the active and resting states over iterations. This uncertainty is also reflected by $\hat{\rho}_t$ as defined in Section 4.1.

Figure 6 presents a summary of group level comparison for various metrics. The consistent pattern across the board indicates that all three metrics can be used to compare different groups, where severity of physical challenges is strongly associated with the amount of activities. Despite the consistent pattern observed in the VM measurements, they can be less informative in practice, as their translation into clinical terms is not as straightforward.

5. Conclusion. We develop a statistical methodology to conduct a comprehensive analysis of data obtained from accelerometer-based wearable devices, taking the advantage of the abundant and detailed information available. A Bayesian approach built on an RJMCMC algorithm is proposed to derive various metrics from the dataset by estimating the unknown latent state of each measurement. The algorithm embodies the alternating nature between the active and resting states of the device wearers. The proposed method is applied to a case study dataset obtained for a clinical study involving subjects with physical impairments. We illustrate several additional metric analyses obtained via our methodology, showcasing the benefit of potential applications. Also proposed is an intuitive and widely applicable framework to aggregate posterior inference from individual subject-level datasets for conducting a post-hoc group-level comparison study. Following the case study, practitioners can conduct various in-depth research focusing on physical activity within our framework. With its capacity to derive various quantities to measure physical activities associated with the data from the wearable devices, we anticipate various subsequent studies from clinical practitioners.

Here is a remark on potential topics for future research. The conversion algorithms used in many devices are often not shared with researchers, leading to potential issues of data inconsistency and lack of reliability due to unknown algorithm changes. Our methodology relies

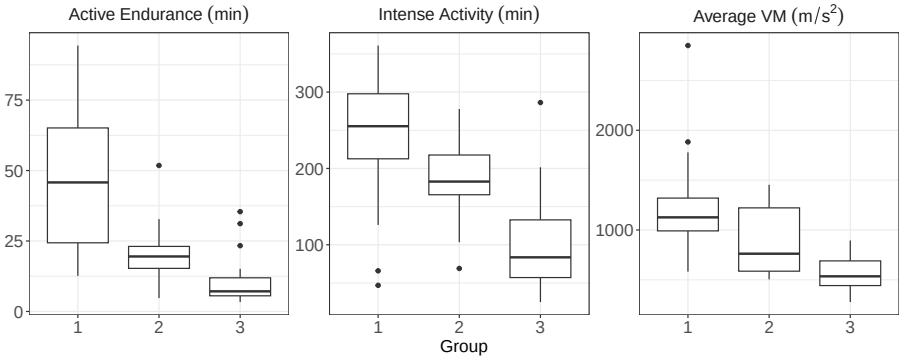

FIG 6. *Box plots are used for group comparison, treating each individual posterior mean as a data point, where three panels depict the active endurance, intense activity, and average VM, respectively.*

solely on raw accelerometer measurements, with a flexibility to produce valuable metrics for practitioners, and it can be extended to handle more complex, larger-scale data. Examples include a model that can adjust differences between multiple devices or multicenters involved in a study.

SUPPLEMENTARY MATERIAL

The Supplementary Material provides the detail of posterior computation in Algorithm 1.

Supplement.pdf
Detail of Posterior Posterior Computation.

Rcode.zip
The `BayesActigraph` package source file and an R Code example to reproduce Table 2, Figure 3, Figure 4, Figure 5, and Figure 6.

REFERENCES

BIANCHIM, M. S., MCNARRY, M. A., BARKER, A. R., WILLIAMS, C. A., DENFORD, S., THIA, L., EVANS, R. and MACKINTOSH, K. A. (2023). A Machine Learning Approach for Physical Activity Recognition in Cystic Fibrosis. *Measurement in Physical Education and Exercise Science* **28** 172–181.

CHOI, L., LIU, Z., MATTHEWS, C. E. and BUCHOWSKI, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and Science in Sports and Exercise* **43** 357–364.

CHOI, L., BECK, C., LIU, Z., MOORE, R., MATTHEWS, C. E. and BUCHOWSKI, M. S. (2021). PhysicalActivity: Process Accelerometer Data for Physical Activity Measurement R package version 0.2-4.

CONRADSSON, D., NERO, H., LÖFGREN, N., HAGSTRÖMER, M. and FRANZÉN, E. (2017). Monitoring training activity during gait-related balance exercise in individuals with Parkinson's disease: a proof-of-concept-study. *BMC Neurology* **17** 1–8.

FAN, Y. and SISSON, S. A. (2011). Reversible jump MCMC. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) 67–92. Chapman and Hall/CRC.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, Third ed. CRC press.

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732.

GREEN, P. J. and MIRA, A. (2001). Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* **88** 1035–1053.

ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96** 161–173.

JOHN, D., MORTON, A., ARGUELLO, D., LYDEN, K. and BASSETT, D. (2018). "What is a step?" Differences in how a step is detected among three popular activity monitors that have impacted physical activity research. *Sensors* **18** 1206.

KIM, Y., BEETS, M. W. and WELK, G. J. (2012). Everything you wanted to know about selecting the "right" Actigraph accelerometer cut-points for youth, but…: a systematic review. *Journal of Science and Medicine in Sport* **15** 311–321.

LAFFERTY, J. D., MCCALLUM, A. and PEREIRA, F. C. N. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001* (C. E. BRODLEY and A. P. DANYLUK, eds.) 282–289. Morgan Kaufmann.

LAWRENCE, G. and MUZA, R. (2018). Assessing the sleeping habits of patients in a sleep disorder centre: a review of sleep diary accuracy. *Journal of Thoracic Disease* **10** S177–S183.

LEAVY, B., HAGSTRÖMER, M., CONRADSSON, D. M. and FRANZÉN, E. (2021). Physical activity and perceived health in people with Parkinson disease during the first wave of COVID-19 pandemic: A cross-sectional study from Sweden. *Journal of Neurologic Physical Therapy* **45** 266–272.

NADER, P. R., BRADLEY, R. H., HOUTS, R. M., MCRITCHIE, S. L. and O'BRIEN, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *The Journal of the American Medical Association* **300** 295–305.

NEAL, R. M. (1999). Regression and Classification Using Gaussian Process Priors. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting June 6-10, 1998* Oxford University Press. https://doi.org/10.1093/oso/9780198504856.003.0021

PALISANO, R., ROSENBAUM, P., WALTER, S., RUSSELL, D., WOOD, E. and GALUPPI, B. (1997). Gross motor function classification system for cerebral palsy. *Developmental Medicine & Child Neurology* **39** 214–23.

RABINER, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77** 257–286.

SUMMERS, G., BOOTH, A., BROOKE-WAVELL, K., BARAMI, T. and CLEMES, S. (2019). Physical activity and sedentary behavior in women with rheumatoid arthritis: a comparison of patients with low and high disease activity and healthy controls. *Open Access Rheumatology: Research and Reviews* **11** 133–142.

TEH, Y. W. (2010). Dirichlet Process. *Encyclopedia of Machine Learning* **1063** 280–287.

TROST, S. G., ZHENG, Y. and WONG, W.-K. (2014). Machine learning for activity recognition: hip versus wrist data. *Physiological Measurement* **35** 2183–2189.

VAN HEES, V. T., SABIA, S., JONES, S. E., WOOD, A. R., ANDERSON, K. N., KIVIMÄKI, M., FRAYLING, T. M., PACK, A. I., BUCAN, M., TRENELL, M. et al. (2018). Estimating sleep parameters using an accelerometer without sleep diary. *Scientific Reports* **8** 12975.

WRIGHT, S. P., HALL BROWN, T. S., COLLIER, S. R. and SANDBERG, K. (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **312** R358–R367.

YOON, M.-J., CHOI, H., KIM, J.-S., LIM, S. H., YOO, Y.-J. and HONG, B. Y. (2022). Physical activity, quality of life and parenting stress in children with cerebral palsy. *Pediatrics International* **64** e15295.