

Original Research

Parametric and Semiparametric Approaches to Analyzing Device-Based Measures of Energy Expenditure in Zucker Diabetic Fatty Rats

Hyunkyung Kim¹, Yuanyuan Luan², Roger S. Zoh², Guoyao Wu³, Carmen D. Tekwe^{2,*}

¹Department of Allergy and Clinical Immunology, Asan Medical Center, 05505 Seoul, Republic of Korea

²Department of Epidemiology and Biostatistics, Indiana University, School of Public Health, Bloomington, IN 47405, USA

³Department of Animal Science, Texas A&M University, College Station, TX 77843, USA

*Correspondence: ctekwe@iu.edu (Carmen D. Tekwe)

Academic Editor: Yudong Cai

Submitted: 30 June 2022 Revised: 28 December 2022 Accepted: 13 January 2023 Published: 20 February 2023

Abstract

Background: Obesity results from a chronic imbalance between energy intake and energy expenditure. Total energy expenditure for all physiological functions combined can be measured approximately by calorimeters. These devices assess energy expenditure frequently (e.g., in 60-second epochs), resulting in massive complex data that are nonlinear functions of time. To reduce the prevalence of obesity, researchers often design targeted therapeutic interventions to increase daily energy expenditure. **Methods:** We analyzed previously collected data on the effects of oral interferon tau supplementation on energy expenditure, as assessed with indirect calorimeters, in an animal model for obesity and type 2 diabetes (Zucker diabetic fatty rats). In our statistical analyses, we compared parametric polynomial mixed effects models and more flexible semiparametric models involving spline regression. **Results:** We found no effect of interferon tau dose (0 vs. 4 $\mu\text{g/kg}$ body weight/day) on energy expenditure. The B-spline semiparametric model of untransformed energy expenditure with a quadratic term for time performed best in terms of the Akaike information criterion value. **Conclusions:** To analyze the effects of interventions on energy expenditure assessed with devices that collect data at frequent intervals, we recommend first summarizing the high dimensional data into epochs of 30 to 60 minutes to reduce noise. We also recommend flexible modeling approaches to account for the nonlinear patterns in such high dimensional functional data. We provide freely available R codes in GitHub.

Keywords: energy expenditure; mixed effects models; spline regression; truncated splines; ZDF

1. Introduction

As sedentary lifestyles spread globally, obesity has become an increasing public health concern [1]. Sedentary lifestyles are growing rapidly with developing technology [2]. Obesity results from a chronic imbalance between food intake and energy expenditure, genetic predisposition, consumption of high fat diets, and inflammation [3]. Additionally, obesity contributes to adverse health outcomes such as insulin resistance, type 2 diabetes, obstructive sleep apnea, osteoarthritis, stroke, hypertension, and cancer [4]. As obesity becomes more prevalent, researchers seek to better understand the causal pathways leading to it. Energy expenditure is a key factor on these pathways and refers to the amount of energy used by the body for all physiological functions, such as movement, respiration, and digestion [5–7]. Energy expenditure has three components: resting metabolism, the thermic effect of feeding, and the thermic effect of physical activity [5,7]. Resting metabolism makes up 60% to 70% of an individual's daily energy expenditure [5]. The thermic effect of feeding, including digestion, accounts for up to 10% of daily energy expenditure [5]. Finally, the thermic effect of physical activity comprises 20% to 30% of daily energy expenditure [5].

Measuring energy expenditure accurately requires sensitive and sophisticated instruments. One commonly used instrument is the open circuit calorimeter, such as the computer-controlled Oxymax metabolic chamber for research animals (Columbus Instruments, Ohio, USA). This instrument measures energy expenditure in epochs of 60 seconds to five minutes during an observation period. The instrument calculates an animal's energy expenditure from its volumetric carbon dioxide production (VCO_2) and volumetric oxygen consumption (VO_2). The device also records the animal's total heat production (heat), and respiratory quotient (RQ). The resulting data are repeated measures that appear as curves or complex high dimensional non-linear functions of time (Fig. 1). Researchers often are confused about the most appropriate method for analyzing these data. A common approach is to compute a summary measure, such as the overall mean energy expenditure for the whole observation period or categorizing individuals on their intensity of activity [8–11]. These approaches are limited because they do not capture variation in energy expenditure or its pattern over time.

Because energy expenditure affects the development of obesity [12–14], some researchers have sought to manipulate energy expenditure and its physiological effects



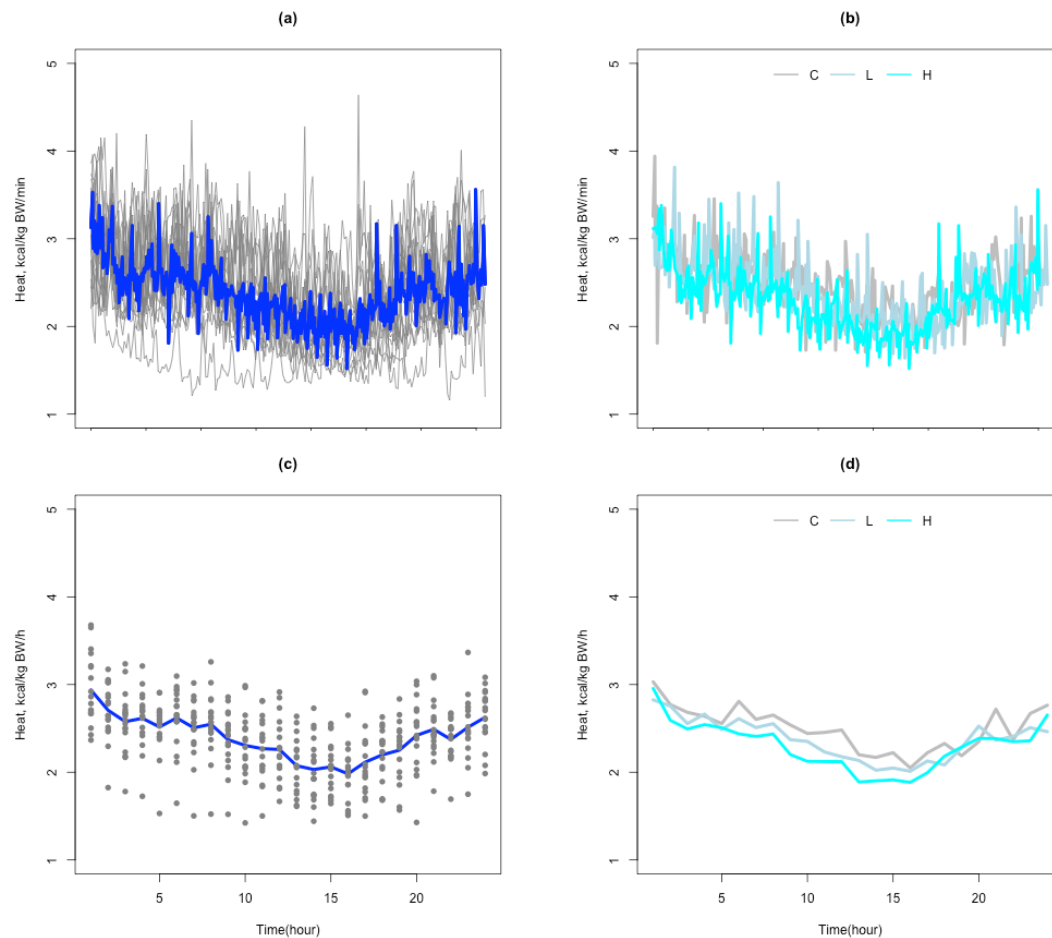


Fig. 1. Plots of the heat production (kcal/kg BW/h) against time (in minutes and hours) over a 24 hour period for the 10-week-old ZDF rats. (a) shows the animal-specific trajectories of untransformed heat production in minutes. (b) shows the animal-specific trajectories of untransformed heat production in minutes by treatment group. (c) shows the animal-specific trajectories of untransformed heat production in hours. (d) shows the animal-specific trajectories of untransformed heat production in hours by treatment group. The blue lines in (a,c) are based on smoothing of the lines. In (b,d), C refers to the control group, L refers to the low dose group, and H refers to the high dose group.

as a way to prevent or reduce obesity. Interferon tau, an anti-inflammatory cytokine, is one proposed intervention for achieving this aim [15,16]. In a previous study, we evaluated the impact of interferon tau on obesity-related outcomes in Zucker diabetic fatty (ZDF) rats [11]. The ZDF animal model has deficiencies in its leptin receptors and therefore researchers often use it for obesity and type 2 diabetes studies. The objective of this study is to provide an introduction to more flexible approaches to assessing intervention effects on high dimensional data frequently collected in biomedical studies such as the device-based measures of energy expenditure.

2. Materials and Methods

We obtained 18 male 23-day-old ZDF rats from Charles River Laboratories and fed them a Purina 5008 diet throughout the study. The Purina 5008 diet consisted of

23.5% crude protein, 6.0% fat, 34.9% starch, 2.6% sucrose, 0.5% glucose plus fructose, 6.8% minerals, and 3.8% fiber, yielding 17,364 kJ gross energy/kg [11]. We kept the study animals in a temperature- and humidity-controlled facility on a 12-h light: 12-h dark cycle. The Texas A&M University Animal Use and Care Committee approved the study (#2010-251).

At 28 days of age, we randomly assigned the rats to receive drinking water (distilled and deionized H₂O) with 0 (control), 4 (low dose), or 8 μ g (high dose) of interferon tau/kg body weight per day (6 rats per condition). The rats had free access to food and drinking water during the 8-week study. To maintain assigned interferon tau dosages, we adjusted concentrations of interferon tau in the drinking water daily based on the volume of water the animals consumed. We changed their drinking water every other day. When the rats were 10 weeks old (week 6 of the inter-

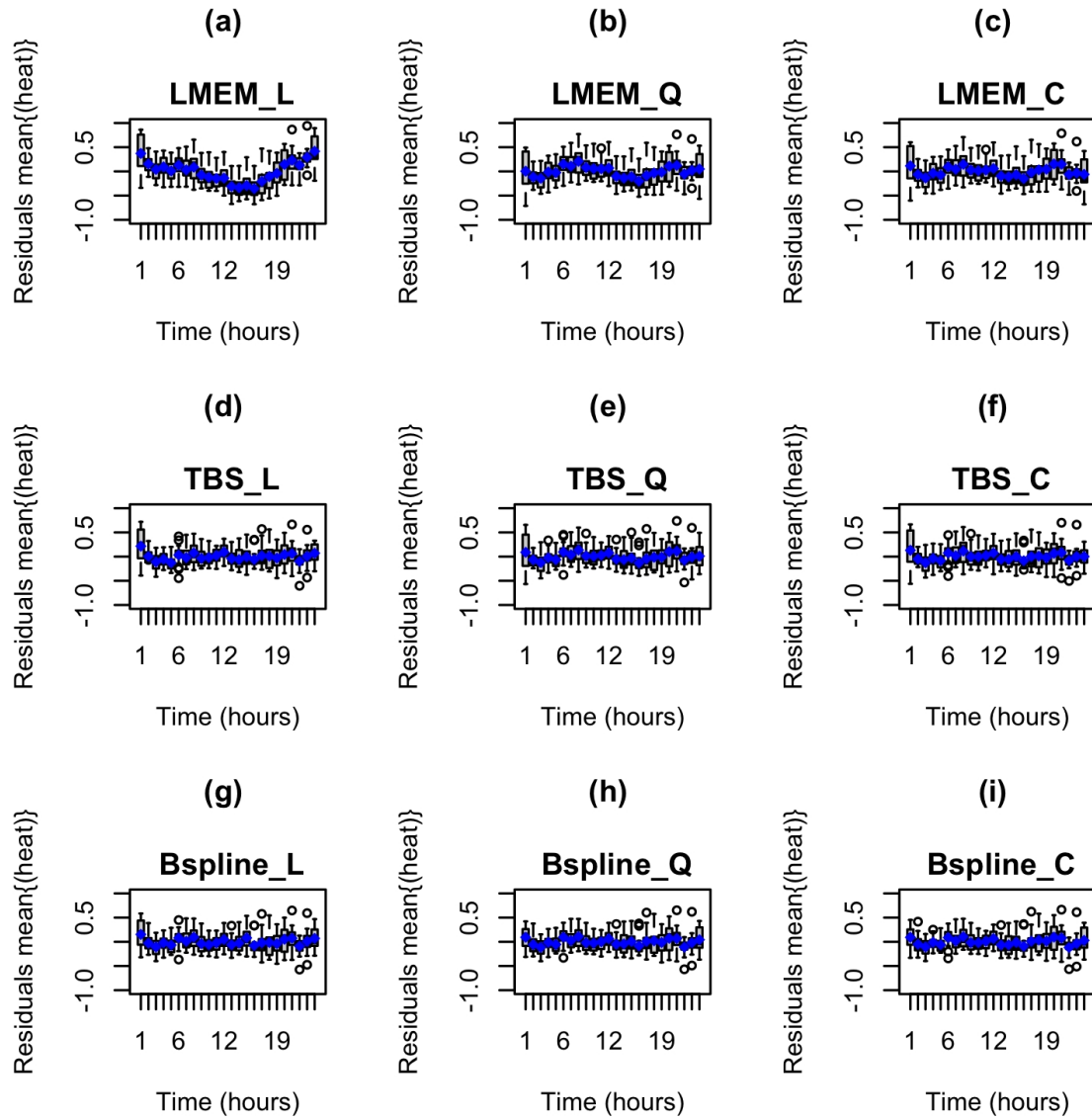


Fig. 2. Boxplots of the residuals for heat production (kcal/kg BW/h) over time (hours) for 10-week-old ZDF rats. (a) LMEM with a linear term for time. (b) LMEM with a quadratic term. (c) LMEM with a cubic term for time. (d) TPBF SMEM with a linear smoothing spline. (e) TPBF SMEM with a quadratic smoothing spline. (f) TPBF SMEM with a cubic smoothing spline. (g) B-spline SMEM with a linear smoothing spline. (h) B spline SMEM with a quadratic smoothing spline. (i) B-spline SMEM with a cubic smoothing spline. The quadratic and cubic terms for time in the LMEMs fit the nonlinear trends in the data better than the LMEM with a linear term for time.

feron tau treatment), we placed each in an Oxymax chamber for 24 hours to assess energy expenditure. Approximately every five minutes, the instrument measured several indicators of energy expenditure: volumetric O₂ consumption (VO₂; L/h/kg body weight [BW]), volumetric CO₂ production (VCO₂; L/h/kg BW), respiratory quotient (RQ; CO₂ production/O₂ consumption) and heat production (kcal/h) (Heat). We focused our analyses on heat production. Our original report has further details on the experiment [11].

3. Models Considered

3.1 Linear Mixed Effects Models

Linear mixed effects models (LMEMs) can be used to analyze repeated measures data [17]. These models extend classical linear regression to correlated data. They provide powerful techniques for analyzing correlated data with complex variance structures, handling missing data, and incorporating nonlinear trends with log or higher order polynomial transformations. LMEMs take the following form:

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_i + \epsilon_{ij}, \quad (1)$$

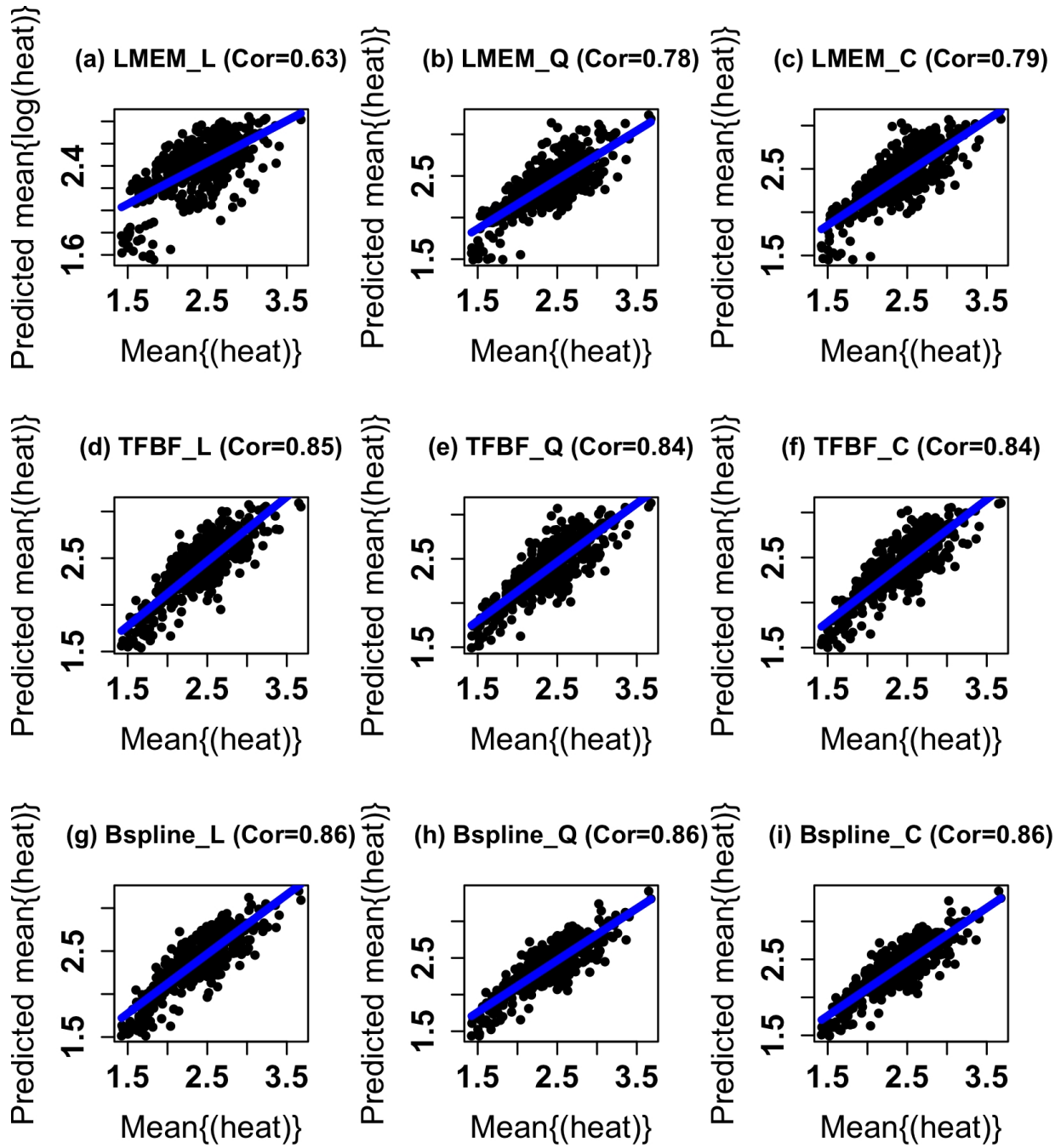


Fig. 3. Predicted values of mean heat production (kcal/kg BW/h) against observed mean heat production(kcal/kg BW/h) in 10-week-old ZDF rats. (a) LMEM with a linear term for time. (b) LMEM with a quadratic term. (c) LMEM with a cubic term for time. (d) TPBF SMEM with a linear smoothing spline. (e) TPBF SMEM with a quadratic smoothing spline. (f) TPBF SMEM with a cubic smoothing spline. (g) B-spline SMEM with a linear smoothing spline. (h) B spline SMEM with a quadratic smoothing spline. (i) B-spline SMEM with a cubic smoothing spline. The quadratic and cubic terms for time in the LMEMs fit the nonlinear trends in the data better than the LMEM with a linear term for time.

where Y_{ij} is the j^{th} response for the i^{th} subject, β is a $p \times 1$ vector of fixed coefficients, X_{ij} is a $1 \times p$ vector of fixed variables, b_i is a $q \times 1$ vector for the random effects, and Z_{ij} is a $1 \times q$ vector for the random variables. The random error terms ϵ_{ij} represent the random variation associated with the

Y_{ij}^{th} response. These models rely on the assumptions that $\epsilon_{ij} \sim Normal(0, \sigma^2)$ and $b_i \sim Normal(0, \Lambda)$, where Λ is the variance-covariance matrix for b_i . The mean response of Y_{ij} is $X_{ij}\beta$, the fixed component of the model, while $Z_{ij}b_i$ is the random component of the model, repre-

Table 1. Results for the LMEMs of heat production (kcal/kg BW/hr) with a linear time term.

Model	Outcome	Variable	$\hat{\beta}$	S.E.	p	AIC
LMEM (raw)	Heat	Intercept	2.589	0.058	<0.0001	4420.8
		L vs. C	-0.141	0.098	0.170	
		H vs. C	0.005	0.098	0.96	
		Time (mins)	-0.0003	0.00	<0.0001	
LMEM (log transformed)	$\log(\text{Heat})$	Intercept	0.933	0.027	<0.0001	-1921.06
		L vs. C	-0.06	0.05	0.21	
		H vs. C	0.011	0.05	0.81	
		Time (mins)	-0.0001	0.00	<0.0001	
LMEM $\text{mean}(\text{Heat})/\text{hr}$	$\text{mean}(\text{Heat})$	Intercept	2.584	0.063	<0.0001	328.575
		L vs. C	-0.141	0.097	0.169	
		H vs. C	0.015	0.097	0.877	
		Time (hr)	-0.015	0.002	<0.0001	
LMEM $\text{mean}\{\log(\text{Heat})\}/\text{hr}$	$\text{mean}\{\log(\text{Heat})\}$	Intercept	0.935	0.029	<0.0001	-398.918
		L vs. C	-0.061	0.045	0.201	
		H vs. C	0.015	0.045	0.749	
		Time (hr)	-0.006	0.001	<0.0001	

p is the parametric p -value obtained based on Z-score, which is given by $Z = \frac{\hat{\beta}}{S.E.}$.

C, control; L, Low dose; H, High dose; hr, hour; min, minutes

senting individual variation from the overall sample mean and allowing description of individual-specific trajectories.

In assessing the impact of oral interferon tau supplementation on energy expenditure, we estimated 12 separate LMEMs resulting from all combinations of energy expenditure transformation (raw or log transformed), unit of time (minutes or hours), and time term (linear, quadratic, or cubic). We intended the log transformations to make the data approximately normal. We performed all analyses using the R Crans Software version 4.2.0 (R Core Team, Vienna, Austria) [18].

3.2 Semiparametric Mixed Effects Models

Penalized spline regression is a flexible semiparametric approach to estimating mean functions in mixed effects models [19]. Mean functions represented by splines can be expressed easily as the best linear unbiased predictors of the mixed effects model [20]. Semiparametric mixed effects models (SMEMs) are also specified as in Eqn. 1. However, the elements of the random components matrices differ from LMEMs. SMEMs include spline basis functions as random effects in addition to subject-specific random effects. Thus, SMEMs can be written as classical mixed effects models that include nonparametric terms for curve smoothing.

We used two kinds of semiparametric functions in our SMEMs: truncated power basis functions (TPBFs) and cubic B-spline functions.

3.2.1 Truncated Power Basis Functions

Truncated power basis functions are simple semiparametric functions that approximate curves. We define a truncated

power function at a given knot κ_k as

$$\begin{cases} (x_i - \kappa_k)^p = 0 & x_i \leq \kappa_k \\ (x_i - \kappa_k)^p > 0 & x_i \geq \kappa_k \end{cases}$$

where p is the order of the polynomial function, and $k = 1, \dots, K$ represents the number of knots [21]. The functions are differentiable up to $p - 1$ times [20–23]. In modeling mean functions, TPBFs approximate curves based on polynomial expansions. A mixed effects model based on truncated power basis is

$$Y_i = X_i\beta + \beta_{0i} + \sum_{k=1}^p \alpha_k (x_i - \kappa_k)_+^p + \epsilon_i \quad (2)$$

where Y_i is the $n_i \times 1$ vector of responses for the i^{th} subject, n_i represents the total number of responses per subject, $X_i\beta$ is the fixed part of the model, and β_{0i} is the subject specific random intercept. The term $(x_i - \kappa_k)^p$ is a p^{th} order truncated power basis of degree p , with κ_k representing the k^{th} knot [20]. Eqn. 2 is a polynomial piece-wise regression model with separate slopes, (α_k) , fit to different partitions of the predictor variable. Thus, $(x_i - \kappa_k)_+$ is an indicator variable indicating the partition where $(x_i - \kappa_k)_+$ is positive. Knots are the points where adjacent partitions meet. For effective estimation, the TPBF approach requires an adequate number of knots or penalization [21].

Our cubic TPBF model of energy expenditure is

$$\begin{aligned} Y_{ij} = & \beta_0 + \beta_1 * \text{time}_i + \beta_2 * \text{time}_i^2 + \beta_3 * \text{time}_i^3 \\ & + \beta_4 * \text{low}_i + \beta_5 * \text{high}_i \\ & + \sum_{i=1}^k \mu_k (\text{time}_i - \kappa_k)_+^3 + b_{0i} + \epsilon_{ij} \end{aligned} \quad (3)$$

Table 2. Results for the LMEMs of heat production (kcal/kg BW/hr) with a quadratic time term.

Model	Outcome	Variable	$\hat{\beta}$	S.E.	p	AIC
LMEM (raw)	Heat	Intercept	2.992	0.06	<0.0001	3819.95
		L vs. C	-0.14	0.098	0.17	
		H vs. C	0.012	0.098	0.91	
		Time (mins)	-0.0002	0.0001	<0.0001	
		Time ²	-0.000	0.000	<0.0001	
LMEM (log transformed)	$\log(\text{Heat})$	Intercept	1.110	0.028	<0.0001	-2554.46
		L vs. C	-0.06	0.05	0.21	
		H vs. C	-0.014	0.05	0.76	
		Time (mins)	-0.0001	0.000	<0.0001	
		Time ²	0.000	0.000	<0.0001	
LMEM $\text{mean}(\text{Heat})/\text{hr}$	$\text{mean}(\text{Heat})$	Intercept	3.052	0.069	<0.0001	174.539
		L vs. C	-0.141	0.097	0.17	
		H vs. C	0.016	0.097	0.87	
		Time (hr)	-0.0124	0.008	<0.0001	
		Time ²	0.004	0.0003	<0.0001	
LMEM $\text{mean}\{\log(\text{Heat})\}/\text{hr}$	$\text{mean}\{\log(\text{Heat})\}$	Intercept	1.133	0.031	<0.0001	-548.795
		L vs. C	-0.061	0.045	0.20	
		H vs. C	0.015	0.045	0.75	
		Time (hr)	-0.052	0.003	<0.0001	
		Time ²	0.002	0.0001	<0.0001	

p is the parametric p -value obtained based on Z-score, which is given by $Z = \frac{\hat{\beta}}{S.E.}$.

Time² is squared time (mins).

C, control; L, Low dose; H, High dose; hr, hour; min, minutes

$\beta_1, \beta_2, \beta_3$ are the fixed coefficients for the linear, quadratic, and cubic terms for time, respectively, and β_4 and β_5 represent the low interferon tau and high interferon tau groups' contrasts, respectively, with the control group. The $(\text{time}_i - \kappa_k)_+^3$ term is the cubic spline basis. We treat the truncated cubic basis splines and the intercept, b_{0i} , as random and assume $\mu_k \sim \text{Normal}(0, \sigma_u^2)$ and $b_{0i} \sim \text{Normal}(0, \sigma_b^2)$. When $\sigma_u^2 = 0$, Eqn. 2 reduces to a mixed effects model. The random effects $(\text{time}_i - \kappa_k)_+^{(p)}$, which we model as normal random curves with mean zero [23], are not present in the LMEM in Eqn. 1. The smoothness of the spline regression rises with increasing degree of the polynomial [23]. The smoothing parameter, $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_\epsilon^2}}$, controls the smoothness of the curve, while the mean square error of the model grows with increasing λ [22,23]. Although easy to construct, models based on the TPBF can be numerically unstable due to correlations between the basis functions. When the range for x_i in Eqn. 2 is wide, the basis functions increase rapidly as x rises. To resolve this issue, the range for x_i may be re-scaled to $[0, 1]$. These disadvantages make the models prone to computational difficulties [21]. B-splines allow analysts to avoid these problems [21,24,25].

3.2.2 B-spline Basis Functions

B-splines allow flexible approaches to analyzing data [21,25]. B-splines are piece-wise polynomial functions of order p connected at their inner knots [19,21,24,26]. While B-splines are equivalent to TPBFs on any given interval $[\kappa_0, \kappa_m]$, they are more numerically stable [20,21,27]. B-splines are transformations of TPBFs [20,21]. To illustrate their equivalence,

let X_T and X_B be design matrices for the TPBF and the B-spline basis functions of the same degree and same knot locations, respectively. Then $X_B = X_T L_p$ where L_p is a square invertible matrix [20].

B-spline basis functions are nonzero over the interval $[k_0, k_{m+1}]$. Next, let $\kappa = (\kappa_0, \kappa_1, \dots, \kappa_m)$ be a set of $m + 1$ non-decreasing knots. The domain for B-splines is $[\kappa_0, \kappa_m]$, with $k_0 = 0$ and $k_m = 1$, typically representing the two boundary knots [24]. We define the k^{th} B-spline basis function of degree p recursively as

$$\begin{cases} B_{k,0}(\kappa) = 1 & \kappa_k \leq \kappa \leq \kappa_{k+1} \\ B_{k,0}(\kappa) = 0 & \text{otherwise} \end{cases}$$

$$B_{k,p}(\kappa) = \frac{\kappa - \kappa_k}{\kappa_{k+p} - \kappa_k} B_{k,p-1}(\kappa) + \frac{\kappa_{k+p+1} - \kappa}{\kappa_{k+p+1} - \kappa_{k+1}} B_{k+1,p-1}(\kappa)$$

In our analyses, we specified the B-spline models as

$$Y_i = X_i \beta + \beta_{0i} + \gamma_i B + \delta B + \epsilon_i \quad (4)$$

where Y_i is the $n_i \times 1$ vector of responses for the i^{th} subject, $X_i \beta$ and δB are the fixed effects, and β_{0i} and γ_i are the subject-specific random intercepts and random slopes for the B-spline basis functions, respectively.

3.3 Inference and Model Selection

One assumption of classical regression models is that covariates are independent. However, polynomial splines in regression models are not independent because they are piece-

Table 3. Results for the LMEMs of heat production (kcal/kg BW/hr) with a cubic time term.

Model	Outcome	Variable	$\hat{\beta}$	S.E.	p	AIC
LMEM (raw)	Heat	Intercept	2.87	0.063	<0.0001	3814.23
		L vs. C	-0.14	0.098	0.173	
		H vs. C	0.014	0.098	0.889	
		Time (mins)	-0.0001	0.0003	<0.0001	
		Time ²	-0.000	0.000	<0.0001	
		Time ³	-0.000	0.000	<0.0001	
LMEM (log transformed)	$\log(\text{Heat})$	Intercept	1.036	0.029	<0.0001	-2583.84
		L vs. C	-0.06	0.05	0.21	
		H vs. C	-0.015	0.05	0.74	
		Time (mins)	-0.0003	0.000	<0.0001	
		Time ²	0.000	0.000	<0.0001	
		Time ³	-0.000	0.000	<0.0001	
LMEM $\text{mean}(\text{Heat})/\text{hr}$	$\text{mean}(\text{Heat})$	Intercept	2.872	0.08	<0.0001	176.975
		L vs. C	-0.141	0.097	0.17	
		H vs. C	0.016	0.097	0.87	
		Time (hr)	-0.045	0.02	<0.0001	
		Time ²	-0.003	0.0002	<0.0001	
		Time ³	-0.0002	0.000	<0.0001	
LMEM $\text{mean}\{\log(\text{Heat})\}/\text{hr}$	$\text{mean}\{\log(\text{Heat})\}$	Intercept	1.045	0.036	<0.0001	-550.11
		L vs. C	-0.061	0.045	0.202	
		H vs. C	0.015	0.045	0.749	
		Time (hr)	-0.014	0.008	0.105	
		Time ²	-0.002	0.0001	0.013	
		Time ³	0.0001	0.000	<0.0001	

p is the parametric p -value obtained based on Z-score, which is given by $Z = \frac{\hat{\beta}}{S.E.}$.

Time² is squared time (mins).

Time³ is cubic time (mins).

C, control; L, Low dose; H, High dose; hr, hour; min, minutes

wise functions used to approximate curves. Therefore, the standard errors and confidence intervals for parameters in classical regression models are not applicable in models involving splines. For inference in spline regression models, nonparametric bootstrap methods can be used [28]. The nonparametric bootstrap involves resampling the data to estimate variances of model parameters without any distributional assumptions. To implement the nonparametric bootstrap, we first resampled the original data with replacement for each animal at different time points in the study. Next, we estimated model coefficients with the resampled data, and then repeated the resampling and estimation process $b = 500$ times. We computed the 95th percent bootstrap confidence intervals using the percentile approach using $[Q_{\alpha/2}, Q_{1-\alpha/2}]$, where $\alpha = 0.05$. The terms $Q_{\alpha/2}$ and $Q_{1-\alpha/2}$ represent the quantiles of the bootstrap distributions for the estimated coefficients.

We also calculated the corresponding p -values for the estimated coefficients under the null hypothesis of $\beta = 0$ as $p\text{value} = 2 * \min [Prob(\hat{\beta} < 0|H_0), Prob(\hat{\beta} > 0|H_0)]$ [28].

We selected models with the smallest Akaike information criteria (AIC) [29,30] values as the best fitting.

4. Results

Summarizing heat production by minute increased variability and random noise in the data relative to summarizing by hour (Fig. 1a–d). Heat production also varied nonlinearly over time. Other device based measures of energy expenditure showed similar patterns as heat production. Therefore, we focus our report on modeling heat production.

4.1 Linear Mixed Effects Models

We estimated twelve separate LMEMs (Tables 1,2,3). The low and high dose groups did not differ significantly from the control group in heat production in any model. Models with a cubic term for time fit the data better than models with a quadratic term, which fit the data better than models with a linear term (see also Figs. 2a–c,3a–c). Also, models with log-transformed heat production fit much better than those with untransformed heat production. Furthermore, models of hourly mean heat production fit better than models of raw heat production at the scale of minutes, although the parameter estimates of paired models (differing only in time units) were very similar. Because coefficients for time terms and their standard errors were often close to the lower bound of zero, inference for these parameters may be inaccurate.

Table 4. Results for TPBF models of heat production (kcal/kg BW/hr) with linear, quadratic, and cubic splines.

Model	Outcome	Variable	$\hat{\beta}$	p^*	CI*	AIC
TPBF (linear spline)	Heat	Intercept	2.883	0.552	(−0.298, 3.247)	3447.53
		L vs. C	−0.108	0.508	(−0.352, 0.256)	
		H vs. C	0.088	0.588	(−0.255, 0.317)	
		Time (mins)	−0.002	0.648	(−0.004, 3.189)	
	<i>mean(Heat)</i>	Intercept	2.706	<0.0001	(2.576, 3.371)	179.369
		L vs. C	−0.106	0.300	(−0.387, 0.148)	
		H vs. C	0.045	0.944	(−0.248, 0.318)	
		Time (hr)	−0.017	<0.0001	(−0.004, −0.001)	
TPBF (quadratic spline)	Heat	Intercept	2.918	0.560	(−0.346, 3.233)	3505.01
		L vs. C	−0.051	0.456	(−0.406, 0.257)	
		H vs. C	−0.010	0.608	(−0.237, 0.314)	
		Time (mins)	−0.002	0.648	(−0.003, 0.000) ¹	
		Time ²	0.000	<0.0001	(0.000, 3.18)	
	<i>mean(Heat)</i>	Intercept	2.981	<0.0001	(2.404, 3.274)	176.249
		L vs. C	−0.106	0.344	(−0.393, 0.253)	
		H vs. C	0.079	0.896	(−0.263, 0.358)	
		Time (hr)	−0.087	<0.0001	(−0.003, −0.001)	
		Time ²	0.003	0.004	(0.000, 0.000) ¹	
TPBF (cubic spline)	Heat	Intercept	1.908	0.520	(−0.322, 3.277)	3601.67
		L vs. C	−0.250	0.512	(−0.393, 0.274)	
		H vs. C	−0.030	0.600	(−0.242, 0.35)	
		Time (mins)	−0.003	0.584	(−0.005, 0.000) ¹	
		Time ²	0.000	0.712	(−0.000, 0.000) ^{1,2}	
		Time ³	0.000	0.900	(−0.000, 3.218) ²	
	<i>mean(Heat)</i>	Intercept	2.743	0.004	(1.738, 3.454)	186.497
		L vs. C	−0.115	0.312	(−0.418, 0.165)	
		H vs. C	0.066	0.988	(−0.275, 0.359)	
		Time (hr)	−0.034	0.008	(−0.006, −0.000) ²	
		Time ²	−0.003	0.216	(−0.000, 0.000) ^{1,2}	
		Time ³	0.000	0.340	(−0.000, 0.000) ^{1,2}	

p^* is the non-parametric p -value calculated as $P = 2 * \min \left[P \left(\hat{\beta} < 0 | H_0 \right), P \left(\hat{\beta} > 0 | H_0 \right) \right]$ under the null hypothesis of $\beta = 0$.

CI* is non-parametric 95% confidence interval obtained by the percentile bootstrap 95% confidence interval.

¹ 0.000 represents small positive values such as 0.000003 that were rounded up to 0.000 when only keeping 3 decimal digits.

² −0.000 represents negative values such as −0.000003 that were rounded up to 0.000 when only keeping 3 decimal digits.

Time² is squared time (mins).

Time³ is cubic time (mins).

C, control; L, Low dose; H, High dose; hr, hour; min, minutes

4.2 Semiparametric Mixed Effects Models

4.2.1 Truncated Power Basis Functions

The TPBF models fit the data substantially better than the LMEMs (Table 4; Figs. 2d–f, 3d–f). The linear spline TPBF model fit raw heat production at the scale of minutes best, while the quadratic spline model fit hourly mean heat production best. As in the LMEMs, there were no statistically significant treatment effects in any TPBF model. Also, TPBF models of hourly mean heat production fit better and had lower AIC values when compared to the AIC values for the analyses conducted at the minute-levels. For the cubic spline models, the higher order terms for time (quadratic and cubic) were not statistically significant. However, both the linear and quadratic terms for time were statistically significant quadratic spline models in the paired models that differed only in time scale.

4.2.2 B-spline Basis Functions

The B-spline SMEMs (Table 5; Figs. 2g–i, 3g–i) fit the data better than the TPBF models and LMEMs. As with all of the other models, there were no statistically significant treatment effects in the B-spline models. Also, B-spline models of hourly mean heat production fit better and had lower standard errors of coefficients than models of raw heat production at the time scale of minutes, although the patterns of coefficients for paired models were similar. The quadratic B-spline analyzed at the hourly level performed the best of all models we estimated for both mean hourly untransformed heat production and untransformed heat production at the time scale of minutes. Time was not statistically significant in the linear spline models for the analyses performed at the hourly and minute levels. However, the linear and quadratic terms for time were statis-

Table 5. Results for B-spline models of heat production (kcal/kg BW/hr) with linear, quadratic and cubicsplines.

Model	Outcome	Variable	$\hat{\beta}$	p^*	CI*	AIC
(SMEM) (linear spline)	Heat	Intercept	2.754	0.560	(−0.292, 3.147)	3379.22
		L vs. C	−0.170	0.596	(−0.325, 0.295)	
		H vs. C	0.037	0.784	(−0.222, 0.309)	
		Time (mins)	0.000	0.648	(−0.001, 3.09)	
	<i>mean(Heat)</i>	Intercept	2.704	<0.0001	(2.604, 3.209)	162.114
		L vs. C	−0.159	0.268	(−0.359, 0.149)	
		H vs. C	0.082	0.744	(−0.246, 0.342)	
		Time (hr)	−0.016	0.004	(−0.001, −0.000) ²	
	Heat	Intercept	3.064	0.568	(−0.272, 3.315)	3344.63
		L vs. C	−0.176	0.620	(−0.344, 0.263)	
		H vs. C	0.093	0.832	(−0.222, 0.319)	
		Time (mins)	−0.002	0.648	(−0.003, 0.000) ¹	
		Time ²	0.000	<00001	(0.000, 3.297) ¹	
(SMEM) (quadratic spline)	<i>meag(Heat)</i>	Intercept	2.899	<0.001	(2.722, 3.364)	142.136
		L vs. C	−0.127	0.316	(−0.367, 0.179)	
		H vs. C	0.100	0.672	(−0.214, 0.36)	
		Time (hr)	−0.115	<0.0001	(−0.003, −0.001)	
		Time ²	0.004	<0.0001	(0.000, 0.0000) ¹	
	Heat	Intercept	2.923	0.552	(−0.276, 3.447)	3376.26
		L vs. C	−0.106	0.636	(−0.327, 0.321)	
		H vs. C	0.065	0.852	(−0.206, 0.36)	
		Time (mins)	−0.003	0.552	(−0.004, 0.000)	
		Time ²	0.000	0.728	(−0.000, 0.000) ^{1,2}	
(SMEM) (cubic spline)	<i>meag(Heat)</i>	Intercept	2.909	<0.0001	(2.741, 3.427)	164.561
		L vs. C	−0.121	0.364	(−0.38, 0.174)	
		H vs. C	0.126	0.620	(−0.224, 0.368)	
		Time (hr)	−0.083	0.028	(−0.004, 0.000)	
		Time ²	0.001	0.472	(−0.000, 0.000) ^{1,2}	
	Heat	Intercept	2.923	0.552	(−0.276, 3.447)	3376.26
		L vs. C	−0.106	0.636	(−0.327, 0.321)	
		H vs. C	0.065	0.852	(−0.206, 0.36)	
		Time (mins)	−0.003	0.552	(−0.004, 0.000)	
		Time ³	0.000	0.640	(−0.000, 3.402) ²	

p^* is the non-parametric p -value calculated as $P = 2 * \min \left[P \left(\hat{\beta} < 0 | H_0 \right), P \left(\hat{\beta} > 0 | H_0 \right) \right]$ under the null hypothesis of $\beta = 0$.

CI* is non-parametric 95% confidence interval obtained by the percentile bootstrap 95% confidence interval.

¹ 0.000 represents small positive values such as 0.000003 that were rounded up to 0.000 when only keeping 3 decimal digits.

² −0.000 represents negative values such as −0.000003 that were rounded up to 0.000 when only keeping 3 decimal digits.

Time² is squared time (mins).

Time³ is cubic time (mins).

(C, control; L, Low dose; H, High dose; hr, hour; min, minutes)

tically significant in the quadratic spline model performed on a the hour level time scale. The quadratic and cubic terms for time were not statistically significant in the cubic spline models.

5. Discussion

Mixed effects models are useful for analyzing repeated measures data. However, with relatively noisy data such as device-based heat production data, variance parameters might not be well estimated. The semiparametric models, especially those with B-splines, approximated the nonlinear patterns in the untransformed heat production data better and thus had substantially higher predictive power than the parametric models. Another advantage of the semiparametric mixed effects modeling approach is that it does not require transforming the

outcome variable (e.g., log transformation of heat production in our LMEMs) to make the data approximately normal and improve model fit. In analyzing energy expenditure data collected by devices, the first step is to evaluate plots of energy expenditure against time in minutes. If there appears to be a considerable amount of random noise in the plots, summarizing the data into longer time periods, such as hours, will reduce the random variation due to the frequency of data collection. If the data represent a high dimensional curve over time, rather than a linear function, we recommend semiparametric mixed effects models with smoothing splines for analysis.

In this manuscript, we demonstrated the use of semi-parametric models to analyze noisy high dimensional data frequently collected by devices in epochs of 60-seconds over multiple days. A common approach to analyzing these data is to

summarize the data into an overall summary such as overall heat production observed over a given week. In our previous analysis of these data [11], we summarized the data to the hourly level and used parametric linear mixed effects models to assess the effects of oral supplementation of interferon tau on device-based measures of energy expenditure. Our analysis included an interaction term between time and the treatment levels. The overall test for the interaction between treatments and time was not statistically significant at the 5% significance level. However, when separate analysis were conducted by each hour of observation to determine the treatment effects at each hour, we observed that the relationship between interferon tau treatment and measures of energy expenditure such as heat production depended on time and that the differences between the animals on the higher doses of interferon and the lower doses depended on time. A limitation of the current study is the sample size. The use of semiparametric methods in assessing treatment effects require larger sample sizes. Our findings have important implications for statistically analyzing data from experimental and clinical studies regarding effects of nutrition (e.g., dietary intakes of amino acids [31]) on improving metabolic profiles and health in animals and humans.

6. Conclusions

With the rise in complex data frequently collected from devices such as the OxyMas instrument, we recommend summarizing the data from units of time in minutes to hourly or half-hourly measures to reduce the noise associated with the frequency of data collection. The use of semiparametric regression methods provide more flexible modeling approaches to analyzing these data compared to parametric methods based on polynomial mixed effects models.

Abbreviations

AIC, Aikake information criteria; LMEM, linear mixed effects model; SMEM, semiparametric mixed effects model; TPBF, truncated power basis function; ZDF, Zucker diabetic fatty.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

HK drafted the manuscript and performed the initial data analysis as part of her master's thesis under the guidance of the last author while at Texas A&M University. YL re-did some of the data analysis and simulations, and assisted with drafting the manuscript. RSZ assisted with developing the concept and data analysis. GW and CDT designed and conducted the animal experiments at Texas A&M University. CDT developed the concept for and assisted with the statistical data analysis, drafting and editing of the manuscript.

Ethics Approval and Consent to Participate

This study (Animal Use Protocol # 2010-251) was approved by The Institutional Animal Care and Use Committee of Texas A&M University. No consent to participate was applicable.

Acknowledgment

The authors would like to thank Dr. Devon Brewer and Dr. Lisa Giles for their assistance in editing this manuscript.

Funding

This research was supported by National Institute of Diabetes and Digestive and Kidney Diseases award number 1R01DK132385-01. The animal experiment of this work was supported by grants from the American Heart Association (10GRNT4480020 and 11GRNT7930004).

Conflict of Interest

The authors declare no conflict of interest. GW is serving as Editorial Board Member of this journal. We declare that GW had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to YC.

Supplementary Material

We provide sample codes for the data analysis, along with instructions on how to use the codes in an Rmd file, in the GitHub repository at https://github.com/ctekwe/EnergyExpenditure_Analysis.

References

- [1] Wyatt SB, Winters KP, Dubbert PM. Overweight and obesity: prevalence, consequences, and causes of a growing public health problem. *The American Journal of the Medical Sciences*. 2006; 331: 166–174.
- [2] Hamilton MT, Hamilton DG, Zderic TW. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. *Diabetes*. 2007; 56: 2655–2667.
- [3] Fernández-Sánchez A, Madrigal-Santillán E, Bautista M, Esquivel-Soto J, Morales-González A, Esquivel-Chirino C, *et al*. Inflammation, oxidative stress, and obesity. *International Journal of Molecular Sciences*. 2011; 12: 3117–3132.
- [4] Abelson P, Kennedy D. The obesity epidemic. *Science*. 2004; 304: 1413.
- [5] Poehlman ET. A review: exercise and its influence on resting energy metabolism in man. *Medicine and Science in Sports and Exercise*. 1989; 21: 515–525.
- [6] Levine JA. Measurement of energy expenditure. *Public Health Nutrition*. 2005; 8: 1123–1132.
- [7] Donahoo WT, Levine JA, Melanson EL. Variability in energy expenditure and its components. *Current Opinion in Clinical Nutrition and Metabolic Care*. 2004; 7: 599–605.
- [8] Tudor-Locke C, Leonardi C, Johnson WD, Katzmarzyk PT, Church TS. Accelerometer steps/day translation of moderate-to-vigorous activity. *Preventive Medicine*. 2011; 53: 31–33.
- [9] Assaad H, Yao K, Tekwe CD, Feng S, Bazer FW, Zhou L, *et al*. Analysis of energy expenditure in diet-induced obese rats. *Frontiers in Bioscience-Landmark*. 2014; 19: 967–985.
- [10] Lamontagne J, Jalbert-Arsenault E, Pepin E, Peyot M, Ruderman NB, Nolan CJ, *et al*. Pioglitazone acutely reduces energy metabolism and insulin secretion in rats. *Diabetes*. 2013; 62: 2122–2129.
- [11] Tekwe CD, Lei J, Yao K, Rezaei R, Li X, Dahanayaka S, *et al*. Oral administration of interferon tau enhances oxidation of energy substrates and reduces adiposity in Zucker diabetic fatty rats. *BioFactors*. 2013; 39: 552–563.

- [12] Storlien LH, James DE, Burleigh KM, Chisholm DJ, Kraegen EW. Fat feeding causes widespread in vivo insulin resistance, decreased energy expenditure, and obesity in rats. *American Journal of Physiology-Endocrinology And Metabolism*. 1986; 251: E576–E583.
- [13] Flatt J. Differences in basal energy expenditure and obesity. *Obesity*. 2007; 15: 2546–2548.
- [14] Foster GD, McGuckin BG. Estimating resting energy expenditure in obesity. *Obesity Research*. 2001; 9: 367S–372S.
- [15] Bazer FW, Spencer TE, Ott TL. Interferon tau: a novel pregnancy recognition signal. *American Journal of Reproductive Immunology*. 1997; 37: 412–420.
- [16] Bazer FW, Ying W, Wang X, Dunlap KA, Zhou B, Johnson GA, *et al*. The many faces of interferon tau. *Amino Acids*. 2015; 47: 449–460.
- [17] Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38: 963–974.
- [18] R Core Team. R: A language and environment for statistical computing. 2022. Available at: <https://www.R-project.org/> (Accessed: 01 February 2022).
- [19] Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 89–102.
- [20] Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press: Cambridge. 2003.
- [21] Eilers PHC, Marx BD. *Splines, knots, and penalties*. Wiley Interdisciplinary Reviews: Computational Statistics. 2010; 2: 637–653.
- [22] Green PJ, Silverman BW. *Nonparametric regression and generalized linear models*. Monographs on Statistics and Applied Probability. 1994; 58.
- [23] Eubank RL. *Nonparametric regression and spline smoothing*. CRC Press: Boca Raton, Florida 1999.
- [24] De Boor C. *A practical guide to splines*. Vol 27. Springer-Verlag: New York. 1978.
- [25] Prautzsch H, Boehm W, Paluszny M. *Bézier and B-spline techniques*. Vol. 6. Springer: Berlin. 2002.
- [26] Dierckx P. *Curve and surface fitting with splines*. Oxford University Press: Oxford. 1995.
- [27] Kagerer K. *A short introduction to splines in least squares regression analysis*. 2013.
- [28] Efron B, Tibshirani RJ. *An introduction to the bootstrap*. CRC Press: Boca Raton, Florida . 1994.
- [29] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19: 716–723.
- [30] Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*. 2012; 17: 228–243.
- [31] Wu G. Amino acids in nutrition, health, and disease. *Frontiers in Bioscience (Landmark edition)*. 2021; 26: 1386–1392.