Author(s): Kortelainen, Lauri; Helske, Jouni; Finni, Taija; Mehtätalo, Lauri; Tikkanen, Olli; Kärkkäinen, Salme

Title: A nonlinear mixed model approach to predict energy expenditure from heart rate

Year: 2021

Version: Accepted version (Final draft)

Please cite the original version:

Kortelainen, L., Helske, J., Finni, T., Mehtätalo, L., Tikkanen, O., & Kärkkäinen, S. (2021). A nonlinear mixed model approach to predict energy expenditure from heart rate. Physiological Measurement, 42(3), Article 035001. https://doi.org/10.1088/1361-6579/abea25

Title: A nonlinear mixed model approach to predict energy expenditure from heart rate

Authors: Kortelainen Lauri[1, 2], Helske Jouni[2], Finni Taija[1], Mehtätalo Lauri[3], Tikkanen Olli[1], Kärkkäinen Salme[2]

Affiliations:

[1]Neuromuscular Research Center, Department of Biology of Physical Activity, University of Jyvaskyla, Finland

[2]Department of Mathematics and Statistics, University of Jyvaskyla, Finland

[3]School of Computing, Faculty of Science and Forestry, University of Eastern Finland

Contact information:

Jouni Helske; e-mail: jouni.helske@iki.fi

Mailing address: Department of Mathematics and Statistics, PO Box 35, FI-40014 University of Jyväskylä, Finland

The study was approved by the ethics committee of the University of Jyväskylä, and followed the declaration of Helsinki and national guidelines for research integrity.

Abstract

Objective. Heart rate (HR) monitoring provides a convenient and inexpensive way to predict energy expenditure (EE) during physical activity. However, there is a lot of variation among individuals in the EE-HR relationship, which should be taken into account in predictions. The objective is to develop a model that allows the prediction of EE based on HR as accurately as possible and allows an improvement of the prediction using calibration measurements from the target individual. *Approach.* We propose a nonlinear (logistic) mixed model for EE and HR measurements and an approach to calibrate the model for a new person who does not belong to the data set used to estimate the model. The calibration utilizes the estimated model parameters and calibration measurements of HR and EE from the person in question. We compare the results of the logistic mixed model with a simpler linear mixed model for which the calibration is easier to perform. *Main results.* We show that the calibration is beneficial already with only one pair of measurements on HR and EE. That is an important benefit over an individual-level model fitting which requires a larger number of measurements. Moreover, we present an algorithm for calculating the confidence and prediction intervals of the calibrated predictions. The analysis was based on up to eleven pairs of EE and HR measurements from each of 54 individuals of a heterogeneous group of people, who performed a maximal treadmill test. *Significance.* The proposed method allows accurate energy expenditure predictions based on only a few calibration measurements from a new individual without access to the original dataset, thus making the approach viable for example on wearable computers.

Keywords: energy expenditure, heart rate monitoring, individual calibration, logistic mixed model, physical activity

Introduction

The importance of physical activity (PA) for humans is generally recognized due to numerous accompanying positive health effects (Biswas et al., 2015; Kesaniemi et al., 2001). Muscle work during physical activity increases energy expenditure (EE), and measuring the PA-induced energy cost is possible with several methods. In laboratory settings, indirect calorimetry is an accurate and broadly applied method where aerobic EE is calculated from oxygen consumption VO_2 measurements (Levine, 2005), whilst in free-living conditions double-labelled water is considered to be the gold standard method (Ainslie et al., 2003; Levine, 2005). However, both of these methods are costly, require expertise, and are impractical to perform outside experimental settings. In comparison, heart rate (HR) monitoring provides a convenient, inexpensive, and practical way to estimate EE using a prediction equation that can be used to approximate the EE during exercise or free living based on an estimated EE-HR relationship. However, there is a lot of variation among individuals in the EE-HR relationship, which needs to be taken into account in model fitting and HR-based prediction of EE.

Traditionally, the relationship between EE and HR is determined for an individual by recording VO_2 and HR simultaneously when the individual performs activities with gradually increasing intensities. It is well-known that at moderate activity levels, EE increases linearly with a positive slope as a function of HR (Livingstone, 1997; Booyens and Hervey, 1960; Oja et al., 1982; Christensen et al., 1983; Haskell et al., 1993). At low levels of activity, the slope is almost zero as the other factors may affect HR without a meaningful change in EE (Ainslie et al., 2003; Achten and Jeukendrup, 2003). To model the discrepancy in slopes, a popular choice has been a piecewise linear model called Flex HR, where EE stays constant at low HR values and increases linearly above a certain HR knot point, which needs to be estimated (Spurr et al., 1988; Livingstone et al., 1990, 2000; Ceesay et al., 1989). A nonlinear model, such as the S-shaped logistic function, provides a smooth curve for the whole range of HR (Li et al., 1993; Moon and Butte, 1996; Dauncey and James, 1979; Schulz et al., 1989; Davidson et al., 1997), which theoretically expresses the relationship between energy expenditure and heart rate quite well, especially at low and moderate intensity levels. It has also been supported empirically (Li et al., 1993; Moon and Butte, 1996; Dauncey and James, 1979).

To date, only *individual-level* nonlinear models for EE have been presented by fitting a logistic model to each individual's data separately. The strength of the individual regression curve approach is the reasonable accuracy of the predictions (Li et al., 1993; Moon and Butte, 1996). However, a lot of measurements per individual are needed, individual prediction curves cannot be used in predicting EE for individuals without measurements, and the variability among individuals is not quantified in such models. In large epidemiological studies, *population-based* prediction equations for EE have been developed (Rennie et al., 2001; Hiilloskorpi et al., 2003; Schrack et al., 2014; Keytel et al., 2005; Charlot et al., 2014). In these works, a linear (mixed) model is used to predict EE using HR and some background variables (e.g. sex, age) as predictors. In mixed models, individual-level random effects are used to model variability among individuals that is not explained by the background variables. This approach allows an accurate prediction of EE for individuals whose measurements were used in the modelling (Schrack et al., 2014), with less measurements needed than for individual-level model. Prediction is possible also for the individuals not included in the original data, but the accuracy can be low because their random effects are unknown.

The nonlinear mixed model of the current work is based on the logistic curve

$$EE = g(\boldsymbol{\phi}, HR) = \frac{\phi_1}{1 + \exp[(\phi_2 - HR)/\phi_3]}, \tag{1}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$ define the S-shape of the curve (Fig. 1). Consequently, the corresponding nonlinear mixed model is called a logistic mixed model. In the mixed model, the three parameters are further written as functions of individual-level random effects and population-level fixed effects related to covariates such as age and sex. The mixed model combines the benefits and overcomes the weaknesses of the population-based and individual-level approaches. The model can be used as a population-based prediction equation by replacing the random effects by their expected value, which is zero. It can be further improved by utilizing measurements of EE and HR from the target individual through random-effect calibration, where the random effects of the target individual are predicted so that the curve moves towards the calibration measurements. The adjustment is the stronger the more measurements are available. The estimated variance-covariance parameters are used in determining the optimal degree of adjustment. No lower limit for the measurements per individual exists; even one measurement can be utilized in an efficient and robust way. Similar approaches have been used in forest sciences (e.g. Lappi, 1991; Hall and Bailey, 2001). In physiology and sports, Schrack et al. (2014) evaluated the calibration using a linear mixed model. However, their approach required that the modelling data set is available for calibration, which is impractical. Our approach requires only that the parameter estimates of the fitted model and calibration measurements from the target individual are available.

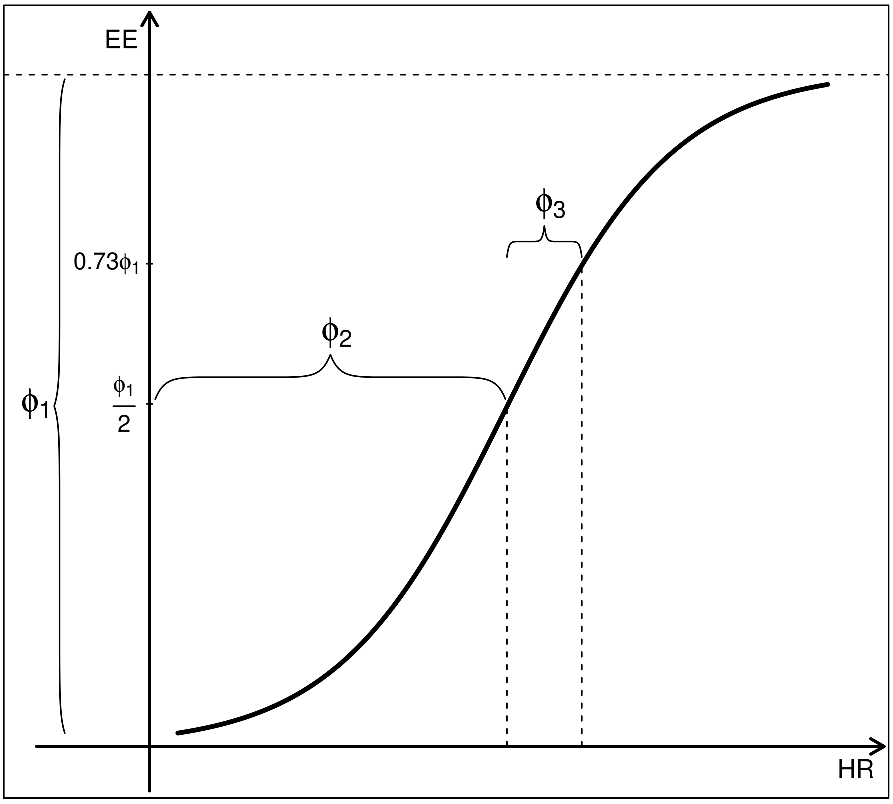

Figure 1. The interpretations of the logistic model parameters demonstrated. A S-shape logistic curve represents EE as a function of HR. A parameter ϕ_1 is the asymptotic maximal energy expenditure, a parameter ϕ_2 is the heart rate value at which EE reaches half of its maximum, and a parameter ϕ_3 is the change in the heart rate when EE increases from the half-maximum to the value of approximately 73% of its maximum.

In this study, the above-described logistic mixed model is fitted and compared with a linear mixed model. A calibration procedure based on a fitted model and measurements of the target individual is presented and its benefit is evaluated with a various number of EE measurements at different calibration loads. An approach to estimate the accuracy of the obtained individual-level curve is also presented. We show that the logistic model provides better predictions than the linear model, and these predictions can be significantly improved with a very small number of calibration measurements from the new individual.

Materials and methods

Data

The data is a part of the EMG24 study "Muscle loading during physical activity and normal daily life: correlates with health and well-being" previously studied in (Tikkanen et al., 2013, 2014; Finni et al., 2016; Tikkanen et al., 2016). The study was approved by the ethics committee of the University of Jyväskylä, and followed the declaration of Helsinki and national guidelines for research integrity. Participants signed a written informed consent prior to the study. All participants of the original dataset are included in our study, but we have omitted some variables not relevant to this study. The description of individuals, test protocol, measurements and processing of data are presented in detail in (Tikkanen et al., 2014) and are repeated here only for the relevant parts. Briefly, a heterogeneous group of 54 individuals carried out a maximal treadmill test with various loads and inclinations. The group consisted of 28 men and 26 women. Mean ± SD of age (yr), BMI ($kg \times m^{-2}$) and Resting HR ($kg \times m^{-2}$) were for men 39.2 ± 13.7, 24.4 ± 3.19 and 53.2 ± 10.5 and for women 39.6 ± 14.2, 22.5 ± 2.46 and 57.8 ±9.39, respectively. The non-standard protocol was chosen to reflect locomotion in free-living conditions better than level loads. In the treadmill test, individuals performed 3-min loads with various walking and running speeds and treadmill inclinations. After rest (load 1), the first six treadmill loads were 2) 4 $km \times h^{-1}$, 3) 5 $km \times h^{-1}$, 4) 5 $km \times h^{-1}$ with 4° descent, 5) 5 $km \times h^{-1}$ with 4° ascent, 6) 6 $km \times h^{-1}$, and 7) 7 $km \times h^{-1}$. With a few exceptions, the individuals under the age of 30 performed load 8) 10 $km \times h^{-1}$ for females and 12 $km \times h^{-1}$ for males. After performing load 9) 5 $km \times h^{-1}$ with 8° ascent for 3 minutes, the maximality of oxygen uptake was evaluated individually. If two out of three of the following criteria were achieved, individuals continued with the same treadmill adjustment (load 10): a) $VO_2max > 85\%$ of the estimated maximum, b) $HR > 90\%$ of the estimated maximum, and c) Borg $RPE > 16$. Otherwise, individuals continued to the load 11) 7 km$\times$h^{-1} with 10° ascent. Individuals performed the last load (10 or 11) until exhaustion.

During the treadmill test, respiratory gases were measured breath-by-breath by with Jaeger Oxycon Pro with the LabManager 3.0 software (Viasys Healthcare Gmbh, Hoechberg, Germany). HR was monitored with a Suunto T6 wrist computer and HR belt (Suunto Oy, Vantaa, Finland). Ventilatory gases and HR were measured constantly and simultaneously during the test. The last 60 seconds at each load were averaged for further analysis. With this averaging we wanted to capture the steady-state values for both heart rate and energy expenditure. We consider that 60 second average from the last minute of each load gives sufficiently accurate presentation of this especially as energy expenditure has some delay in its measurement compared to fast responding heart rate (Hiilloskorpi et al., 2003; Schrack et al., 2014; Keytel et al., 2005). EE was estimated from oxygen consumption (VO_2) and respiratory exchange ratio (RER) with the equation (Lusk, 1924):

$$\text{EE } (kcal/min) = (1.2 \times \text{RER} + 3.85)(\text{VO}_2/1000).$$

The measurements of EE and HR variables used in the analysis for all 54 persons are illustrated in Fig. 2.

Linear and logistic mixed model for EE prediction

As the data included up to eleven repeated measurements from each individual, a mixed model approach was chosen to take account of the inter-individual variation. Both linear and nonlinear (logistic) mixed models were fitted to the data. Also, the calibration of both models was performed.

A linear mixed model with EE as a response variable can be written for an individual *i* as:

$$\mathbf{EE}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i), \quad \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D}),$$

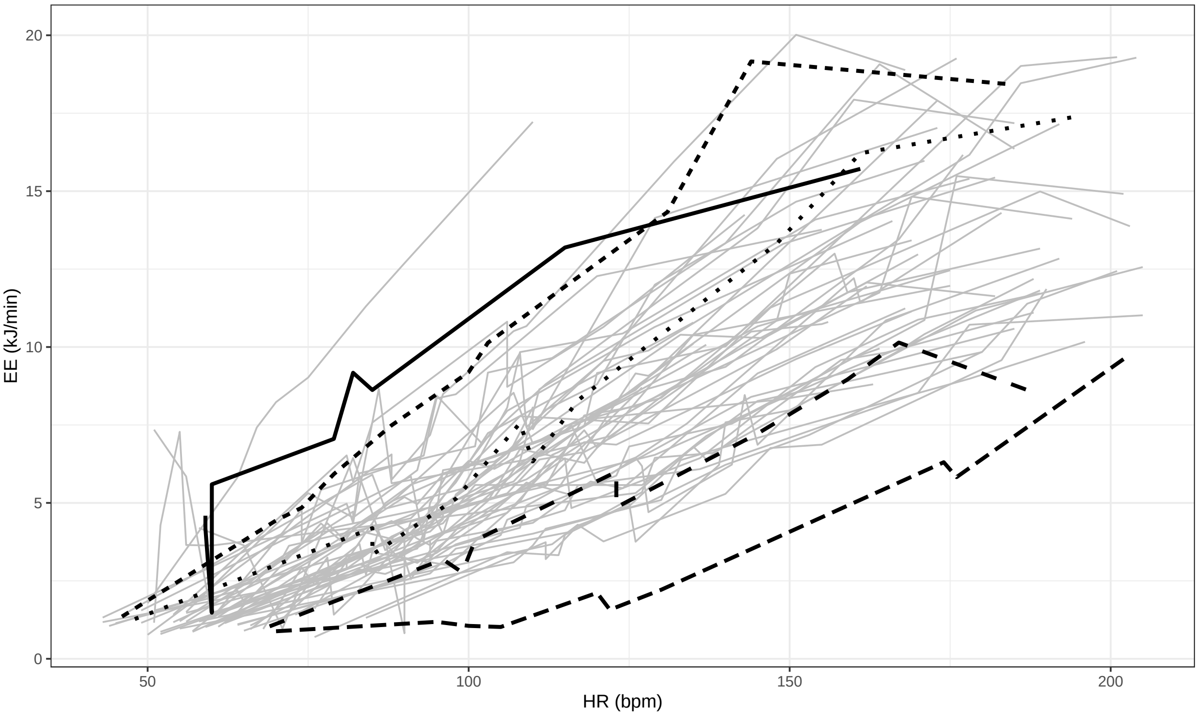

Figure 2. A plot of individual measurements of energy expenditure (EE) and heart rate (HR) with few randomly samples individuals highlighted.

where the error term $\boldsymbol{\epsilon}_i$ and the random effects $\mathbf{u}_i$ are defined to be mutually independent. Here $\mathbf{R}_i$ and $\mathbf{D}$ are positive-definite covariance matrices of $\boldsymbol{\epsilon}_i$ and $\mathbf{u}_i$, respectively. The parameters of both matrices are denoted by an parameter vector $\boldsymbol{\theta}$ which is unknown. Moreover, in our study, the fixed part $\mathbf{X}_i\boldsymbol{\beta}$ contains HR and some of the candidate background variables (e.g. sex and age). The fixed effects $\boldsymbol{\beta}$ are common to the population. Some covariate effects on EE may vary across individuals which is represented as a random part $\mathbf{Z}_i\mathbf{u}_i$ of the model.

The logistic mixed model applied to the EE in relation to HR can be presented for an individual i as follows

$$EE_{ij} = g(\boldsymbol{\phi_i}, HR_{ij}) + \epsilon_{ij} = \frac{\phi_{1i}}{1 + \exp[(\phi_{2i} - HR_{ij})/\phi_{3i}]} + \epsilon_{ij}, \tag{2}$$

where shape parameters $\phi_i = (\phi_{1i}, \phi_{2i}, \phi_{3i})$ contain both fixed and random effects for $k = 1, 2, 3$,

$$\phi_{ki} = \mathbf{a}'_{ki}\boldsymbol{\beta}_k + \mathbf{u}_{ki}. \tag{3}$$

Above, the errors ϵ_{ij} and random effects $\mathbf{u}_{ki}$ follow the normal distribution as in a linear mixed model. The shape parameters have a linear connection to covariate vectors $\mathbf{a}_{ki}$ assembled in matrix $\mathbf{A}_i$, the regression coefficients $\boldsymbol{\beta}$ and the random effects $\mathbf{u}_i$, see Lindstrom and Bates (1990) for details. Note that the linear submodels for the three logistic function parameters are estimated jointly, which ensures efficient use of the data and allows realistic modelling of the joint behavior of the submodels through random effects, which can be correlated across models.

The expected value of EE made with the model based on the logistic curve cannot be negative which is an improvement over the traditional linear model. At low and high EE levels, the slope of the logistic curve is almost horizontal, whereas at moderate activity levels the relation is linear and positive. The shape of the curve and interpretations of the model parameters are illustrated in Fig. 1. All the parameters have physical interpretations: A parameter ϕ_1 is the asymptotic maximal aerobic energy expenditure, a parameter ϕ_2 is the heart rate value at which EE reaches half of its maximum, and a parameter ϕ_3 is the change in the heart rate when EE increases from the half-maximum to the value of approximately 73% of its maximum (Pinheiro and Bates, 2000).

Model selection based on RMSE

As we aimed at good out-of-sample predictive accuracy, we used cross-validation based on the root mean square error (RMSE) as a measure for selecting an optimal set of background covariates for the models (see, e.g, Hastie et al. (2009, Chapter 7.10) for details on cross-validation).

As potential predictors in the candidate models, we used HR (always present), age, sex, height, weight, BMI, and resting HR, which are easy and inexpensive to measure. For a linear mixed model, we considered models with HR and all main effects and pairwise interactions with HR, with an individual-level random intercept and regression coefficient for HR. For a logistic mixed model, in order to keep the search-space of potential models reasonable, we considered only the main effects of all predictors for each of the three shape parameters. For random effects of the logistic model, we used only intercepts and tested different correlation structures for these (i.e., a full covariance matrix of three intercept terms, all combinations of a single pairwise correlation, and fully uncorrelated intercepts). For both linear and logistic models, we allowed the residual variance to depend on sex. The parameters were estimated by

REML (restricted maximum likelihood) with the Lindstrom-Bates algorithm (Lindstrom and Bates, 1990) by using functions `lme` and `nlme` of R (R Core Team, 2020) package `nlme` (Pinheiro et al., 2020) for linear and logistic models, respectively.

For the variable selection, we used leave-one-individual-out cross-validation as follows. Given a candidate model, we estimated the model parameters for each subset of data where one individual was left out, and then computed the RMSE for the leave-out individual using the prediction based on fixed part of the model. While in theory predictions at the population-level using nonlinear mixed models such as logistic mixed models requires integrating over the random effects of the mixed model, here we followed the common approach where population-level predictions are computed by zeroing-out the random effects. The overall cross-validation error for the candidate model was then obtained by averaging these RMSE values over all individuals, and the model with the lowest average RMSE was selected for further analysis. This approach differs from the one used in Tikkanen et al. (2014), where the selection of the background variables age and sex was based on statistical significance (no other variables than age and sex were tested).

Calibration of prediction equations for new individuals when original data not available

The idea behind the calibration of a mixed model is to provide improved predictions for individuals who have not been a part of the original data used for fitting the prediction model. Without having any information from a new individual's EE values, a general prediction can be made using the fixed part of the mixed model. However, if simultaneous measurements on EE and HR can be collected similarly as for the original data, predictions of EE can be improved by using these measurements and parameter estimates for the original model. When there are only few measurements from the new individual, the mixed model borrows strength from other participants' measurements by shrinking the predictions towards the population average, an effect which decreases when the number of measurements of the new individual is increased.

In calibration the random effects are predicted for a new individual. In the linear case, this is relatively straightforward. The prediction of $\mathbf{u}$ is calculated as when fitting a linear mixed model (Laird and Ware, 1982), except for the fact that the new measurements do not contribute to original parameter estimates. In the nonlinear case, the procedure is more complicated, and the prediction of $\mathbf{u}$ must be done numerically (see Appendix A). The iterative algorithm used for the calibration of the nonlinear model was first introduced by Meng and Huang (2009) in the forestry context, see Mehtätalo and Lappi (2020) for discussion and examples. The calibration code for the logistic mixed model, written in R language (R Core Team, 2020), can be found in supplementary materials.

Calibration protocol and evaluation

Estimated linear and logistic mixed models were calibrated for a new individual as follows: Since all individuals performed the loads 1, 3, 5 and 7, they were used as calibration loads. The rest of the loads were left as a test set. As for the model selection, we used the cross-validation approach to evaluate the benefits of the calibration for the EE prediction. First, one individual's measurements were excluded from the data and a model was fitted to this partial data consisting of the rest 53 individuals' measurements. Second, the prediction model was calibrated for the left-out individual using the chosen calibration load(s). Third, the EE predictions were performed with this calibrated model using the predictor values of the individual's test set loads, and the corresponding RMSE was recorded. The process was repeated for each individual, which gave us the average RMSE for the particular calibration load(s). This procedure was then repeated for each combination of calibration loads.

For illustrating the uncertainty accompanying the predictions, using logistic and linear mixed models with and without calibration, a parametric bootstrap method (Efron and Tibshirani, 1994) was used (see Appendix B) to compute prediction curves and intervals for an individual who benefitted from the calibration.

Results

Using the variable selection based on cross validation, the final model equations for the logistic and linear mixed models are presented in Table 1 and the corresponding estimated parameters and their confidence intervals in Table 2. In addition to HR, candidate covariates for the models were age, sex, height, weight, BMI and resting HR, which are easy and inexpensive to measure. Before estimating the final models, the covariates weight, resting HR and age were centred in 70 kg, 56 bpm and 39 years, respectively. The final logistic mixed model included, in addition to individual random intercept terms, covariates sex and weight, resting HR, and age, which were used in modelling parameters ϕ_1, ϕ_2, and ϕ_3, respectively. According to RMSEs for the model selection, a correlation parameter between the random intercepts of ϕ_2 and ϕ_3 was also required. In the case of the linear mixed model, the final model used sex, resting HR, weight and HR, as well as the interaction of weight and HR as fixed effects.

Variables that ended up in the logistic model seem rather logical: the sub-model for ϕ_1 includes weight and sex which are highly correlated with the energy expenditure of an individual, whilst the sub-model for ϕ_2 includes resting HR, which is rather variable between individuals and mainly dependent on fitness level of an individual. Fitness level also affects the maximal energy expenditure of an individual, so for individuals with higher fitness levels the whole 'heart rate – energy expediture' curve is more to the left and up compared to individuals with lower fitness level. Furthermore, the sub-model for ϕ_3 includes age, which is associated with decreased maximum heart rate. Thus, for older individuals the whole 'heart rate – energy expediture' curve is more down compared to younger individuals thus making it rather logical that age is in the ϕ_3 part of the equation.

TABLE 1. The model equations for the logistic and the linear mixed model.

The logistic mixed model

$$\phi_{1i} = (\beta_{10} + u_{1i}) + \beta_{11} \times Male_i + \beta_{12} \times Weight_i$$

$$\phi_{2i} = (\beta_{20} + u_{2i}) + \beta_{21} \times RestHR_i$$

$$\phi_{3i} = (\beta_{30} + u_{3i}) + \beta_{31} \times Age_i$$

$$EE_{ij} = g(\boldsymbol{\phi}_i, HR_{ij}) + \epsilon_{ij}$$

The linear mixed model

$$EE_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) \times HR_{ij} + \beta_2 \times Male_i + \beta_3 \times RestHR_i + \beta_4 \times Weight_i + \beta_5 \times HR_{ij}Weight_i + \epsilon_{ij}$$

In the model equations, βs and us represent the fixed and the random effects, respectively.
A dummy variable $Male_i$ has a value 1 if the individual *i* is male, otherwise 0.
A residual $\epsilon_{ij} \sim N(0, \sigma^2\gamma^2_{male_i})$ where γ_{male_i} is estimated if the individual is male, otherwise 1.
Using average values, $Weight_i$ was centred by 70 kg, $RestHR_i$ by 56 bpm, and Age_i by 39 years.

TABLE 2. The parameter estimates for the logistic and linear mixed models.

Fixed part for the logistic model		
ϕ_1	**Estimate**	**95% CI**
β_{10}	14.90	(13.99, 15.82)
β_{11}	1.87	(0.53, 3.22)
β_{12}	0.18	(0.12, 0.23)
ϕ_2		
β_{20}	122.08	(117.27, 126.89)
β_{21}	0.45	(0.10, 0.79)
ϕ_3		
β_{30}	28.97	(26.64, 31.30)
β_{31}	0.09	(-0.03, 0.20)
Random part		
SD(u_{1i})	0.90	(0.34, 2.35)
SD(u_{2i})	14.10	(11.21, 17.73)
SD(u_{3i})	5.39	(3.89, 7.48)
corr(u_{2i}, u_{3i})	0.54	(0.20, 0.77)
Residual		
σ	0.84	(0.76, 0.93)
γ_{male_i}	1.35	(1.17, 1.56)

σ and γ_{male_i} are the parameter estimates of the variance function.
γ_{male_i} takes the estimated value if the individual *i* is male, 1 otherwise.

Fixed part for the linear model		
	Estimate	**95% CI**
β_0	-5.33	(-5.95, -4.71)
β_1	0.11	(0.10, 0.11)
β_2	0.49	(-0.35, 1.34)
β_3	-0.06	(-0.09, -0.02)
β_4	-0.06	(-0.11, -0.02)
β_5	1.5×10^{-3}	$(1.1 \times 10^{-3}, 2.0 \times 10^{-3})$
Random part		
SD(u_{0i})	1.38	(0.98, 1.94)
SD(u_{1i})	0.02	(0.01, 0.02)
corr(u_{0i}, u_{1i})	-0.67	(-0.84, -0.37)
Residual		
σ	0.93	(0.85, 1.03)
γ_{male_i}	1.40	(1.22, 1.61)

σ and γ_{male_i} as in the logistic model.

More precisely, according to the logistic mixed model, weight has a positive effect on maximal EE and men had on average higher maximal EE than women (ϕ_1) (Table 1). Resting HR had a positive effect on the half of the maximum EE. Age had a positive impact on the change in heart rate when EE increases from the half-maximum to the value of approximately 73 % of its maximum (ϕ_3). In the case of a linear mixed model, men had on average higher EE than women, high resting HR decreased EE, HR had a positive effect, weight negative, and interaction term between HR and weight positive. The model assumptions about expected value, variance and normality, were fulfilled quite well both in linear and nonlinear cases.

When the fixed part of the logistic mixed model was used for population-level prediction, i.e. without predicted random effects, the RMSE of prediction was 1.67 kcal/min (Table 3). As expected, the individual-level predictions based on random-effect calibration were better, with up to 40% decrease in average RMSE compared to the population-level prediction. The single most beneficial calibration load was the load 7, which alone improved the results almost as much as all four loads. The rest load did not improve the prediction accuracy when used in calibration. The logistic mixed model outperformed the linear mixed model in every case.

TABLE 3. Calculated average (over participants) RMSEs and their standard errors for different combinations of calibration loads when using the calibrated linear and logistic models.

	Rest	Load 3	Load 5	Load 7	Logistic RMSE (kcal/min)	Linear RMSE (kcal/min)
1					1.67 (0.13)	1.83 (0.14)
2	X				1.69 (0.13)	1.82 (0.14)
3		X			1.27 (0.08)	1.41 (0.09)
4			X		1.41 (0.15)	1.46 (0.10)
5				X	1.10 (0.07)	1.30 (0.09)
6	X	X			1.32 (0.08)	1.45 (0.10)
7	X		X		1.31 (0.10)	1.44 (0.09)
8	X			X	1.13 (0.07)	1.34 (0.09)
9		X	X		1.25 (0.10)	1.34 (0.09)
10		X		X	1.01 (0.07)	1.21 (0.08)
11			X	X	1.05 (0.07)	1.23 (0.08)
12	X	X	X		1.22 (0.09)	1.35 (0.09)
13	X	X		X	1.03 (0.07)	1.26 (0.08)
14	X		X	X	1.06 (0.07)	1.26 (0.08)
15		X	X	X	1.02 (0.07)	1.19 (0.07)
16	X	X	X	X	1.01 (0.07)	1.21 (0.08)

We now illustrate the practical application of the proposed method in case of the logistic mixed model. The general workflow is as follows. Note that first two steps are only performed once using the original study data, and steps (3)-(5) are performed when calibrated predictions for a new individual are needed.

1. Using the original data, estimate and store the parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ of the logistic model defined in Table 1.

2. Optionally, if there is a need for prediction and/or confidence intervals, obtain and store bootstrap samples of the model parameters using steps (1)–(4) of the bootstrap algorithm of Appendix B.

3. With measured background covariates and at least one measurement pair of EE and HR for a new target individual i, estimate the individual's random effects $\hat{\mathbf{u}}_i$ using the parameters from step (1) and the calibration algorithm of Appendix A.

4. Predict EE values for the new individuals at given HR values using the logistic function $g(\mathbf{A}_i\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_i, \mathbf{HR})$.

5. Optionally, using the bootstrap samples from step (2), compute prediction and/or confidence intervals using steps (5)–(10) of the bootstrap algorithm of Appendix B.

Fig. 3 shows 90 % prediction intervals, with and without calibration using both linear and logistic model, for an example individual (the male of age 28, weighting 54 kilograms with resting HR 88) who benefited from the calibration. The confidence and prediction intervals were much narrower when calibration was performed. This happens because majority of the uncertainty in the population-level predictions is attributed to the variability between individuals, which is remarkably decreased by the prediction of the random effects using the calibration measurements.

Discussion

The purpose of this study was to show the feasibility of using the logistic mixed model in representing the HR effect in relation to EE and of using calibration of the model for a new

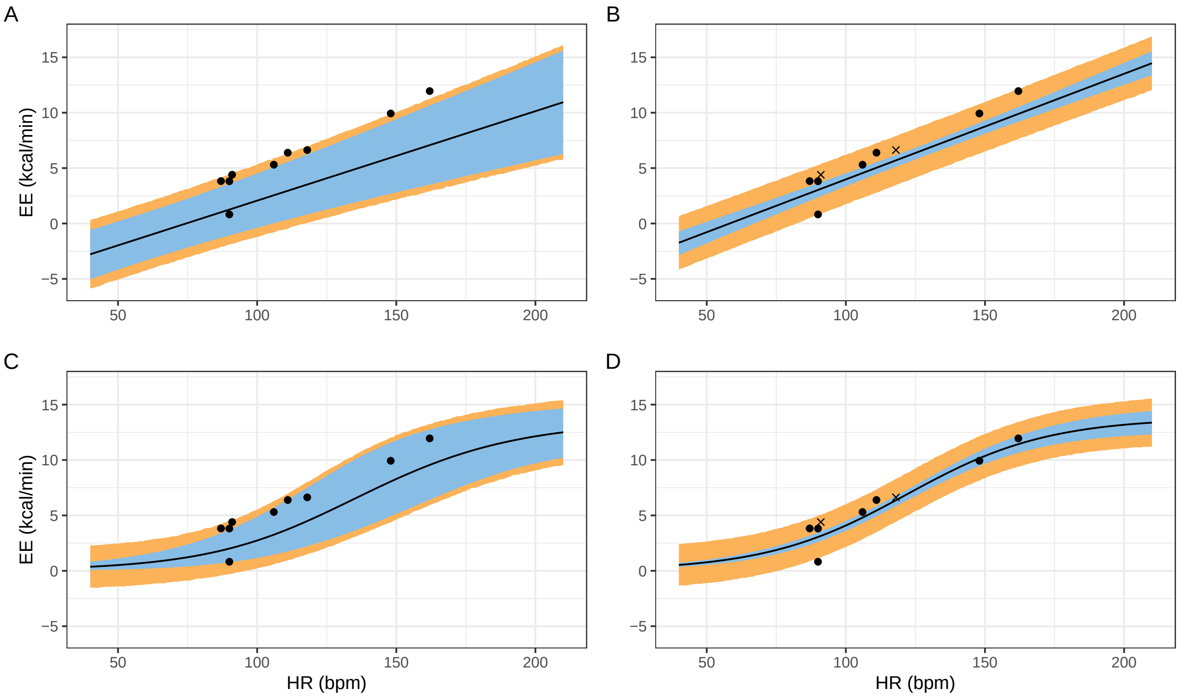

Figure 3. Prediction and confidence intervals for an individual without (left panel) and with calibration (right panel), when using a linear mixed model (A and B) and a logistic mixed model (C and D), respectively. The inner band seen on the figures represents the 90 % confidence interval for the prediction and the outer band is the prediction interval for a new measurement of EE. The used calibration loads 3 and 7 are marked with a cross in the figures, while the individual's other measurements are marked with black dots. Additionally, the solid line represents the point estimates (the mean of predictions).

individual to increase the accuracy of the prediction. We found that the logistic mixed model performed better than the linear mixed model when predicting EE, both at population-level and with calibration. In addition, the accuracy of the EE predictions can be improved remarkably through individual calibration even when based only on one load (Table 3) and the related prediction interval for an individual-level EE-HR curve is remarkably narrower (Fig. 3, orange bands) compared with the population-based EE-HR curve.

Several population-based EE prediction equations with HR have been presented in the past, however, none of them are nonlinear (Hiilloskorpi et al., 2003; Schrack et al., 2014; Keytel et al., 2005; Charlot et al., 2014). Some researchers (Li et al., 1993; Moon and Butte, 1996; Dauncey and James, 1979) have used the nonlinear, logistic model in modelling EE vs HR relation but fitting every individual's data separately. Li et al. (1993) concluded that the individual EE vs HR curves may differ vastly within the same individual at different times. The mixed model approach may overcome this difficulty to some extent as coefficients of the fixed part are determined by all data and changes in the predictor variables (e.g. age, weight and resting HR) also change the prediction equation. In our study, the estimated models included weight as a fixed effect, hence the changes in weight within the individual can be taken into account in the curve shape. In the same way, the effect of ageing on maximal EE is considered in the model. As observed in several studies (e.g. Spurr et al., 1988; Livingstone et al., 1990; Luke et al., 1997), EE is linearly related to HR except for the lowest and the highest level of physical stress in which cases EE vs HR curve is almost horizontal. Factors independent of physical activity (e.g. caffeine, insufficient sleep, humidity, smoking, emotional stress) may affect HR at rest or under low physical strain. Also in our study the largest discrepancies between observed values of EE and predictions of EE were at low HR values: As shown the left-hand side of Fig. 2, there were few cases with relatively large EE values compared to small HR values. However, the majority of data followed the logistic model well.

In this study, we applied the statistical method called the calibration of mixed models to determine how measurements of a new individual (not included in the original data) enhance the prediction accuracy of EE. The method is the most useful when collecting new data is time-consuming and/or expensive but still some increment in accuracy is desired. The method is extensively applied in forest sciences for instance in the estimation of tree heights for a given forest stand (Mehtätalo and Lappi, 2020).

The results of calibration (Table 3) shows the measurements taken at rest will not improve the prediction accuracy. This is not surprising, as mentioned above, factors other than physical activity may have an impact on HR at rest, and thus, the measurements taken at rest will not be informative when the physical strain increases. It can also be seen in Table 3 that the logistic model outperforms the linear model both at the population-level and at every level of calibration. Again, this was anticipated as the logistic model is well-supported physiologically at low and moderate activity levels. At the high activity level, the energy expenditure could theoretically increase instead of being horizontal as in logistic curve (McArdle et al., 2015). However, with our data, the logistic model seems to work quite well also at high levels, when few calibration points at a low or moderate level is used.

According to the results, the most beneficial calibration load was load 7 (7 km$\times$h^{-1}) which was the most demanding among the calibration loads. This raises the question if using even more demanding loads in calibration could be worthwhile. An experimental design to study this question would be more time-consuming as breaks between performing the loads would be needed. Furthermore, performing the more demanding loads without carrying out the less demanding ones first could increase the risk for injury.

Calculating the prediction intervals was done numerically with using a parametric bootstrap method. It can be seen in Fig. 3b that the precision of predictions was increased near the observations used for calibration. The figure also shows how the calibrated curves (the linear and logistic) are shifted downwards giving a better fit for the data and eventually more accurate predictions of EE.

The introduced method for predicting EE by HR is a combination of the individual regression approach and the population-based approach. If there are no measurements available for an individual whose EE predictions are of interest, the predictions can be performed with the fixed part of the model, which provides a population-based prediction equation. On the other hand, if improved accuracy is desired but resources are limited, a single pair of simultaneous measurements of EE and HR could lead to a sufficient improvement in accuracy.

In conclusion, the developed nonlinear logistic mixed model and its calibration provides a superior estimation method for energy expenditure based on HR recordings, compared with linear mixed models, especially during activities of daily life that typically are submaximal in nature and occur in varying terrain.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

The present study was supported by the Academy of Finland (#128643/T.Finni) and Ministry of Education and Culture (42/627/2010/ and 59/626/2016/T.Finni). Jouni Helske was supported by the Academy of Finland research grant 311877. **Author contributions**: TF acquired funding for the study. TF and OT designed the experiment and SK, LM, LK and JH designed this particular study. OT collected and processed the data. JH, LM, SK and LK planned and JH and LK performed the statistical analyses. LK, JH, TF, LM, OT, and SK drafted and were involved

in interpretation of the results, provided critical comments on the content of the manuscript and gave final approval for the manuscript.

References

Achten, J., and Jeukendrup, A.E. 2003. Heart rate monitoring. Sports Medicine 33: 517–538.

Ainslie, P.N., Reilly, T., and Westerterp, K.R. 2003. Estimating human energy expenditure. Sports Medicine 33: 683–698.

Biswas, A., Oh, P.I., Faulkner, G.E., Bajaj, R.R., Silver, M.A., Mitchell, M.S., and Alter, D.A. 2015. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: A systematic review and meta-analysis. Annals of Internal Medicine 162: 123–132.

Booyens, J., and Hervey, G. 1960. The pulse rate as a means of measuring metabolic rate in man. Canadian Journal of Biochemistry and Physiology 38: 1301–1309.

Ceesay, S.M., Prentice, A.M., Day, K.C., Murgatroyd, P.R., Goldberg, G.R., Scott, W., and Spurr, G. 1989. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. British Journal of Nutrition 61: 175–186.

Charlot, K., Cornolo, J., Borne, R., Brugniaux, J.V., Richalet, J.P., Chapelot, D., and Pichon, A. 2014. Improvement of energy expenditure prediction from heart rate during running. Physiological Measurement 35: 253.

Christensen, C.C., Frey, H.M., Foenstelien, E., Aadland, E., and Refsum, H.E. 1983. A critical evaluation of energy expenditure estimates based on individual O2 consumption/heart rate curves and average daily heart rate. The American Journal of Clinical Nutrition 37: 468–472.

Dauncey, M., and James, W. 1979. Assessment of the heart-rate method for determining energy expenditure in man, using a whole-body calorimeter. British Journal of Nutrition 42: 1–13.

Davidson, L., McNeill, G., Haggarty, P., Smith, J.S., and Franklin, M.F. 1997. Free-living energy expenditure of adult men assessed by continuous heart-rate monitoring and doubly-labelled water. British Journal of Nutrition 78: 695–708.

Efron, B., and Tibshirani, R.J. 1994. An Introduction to the Bootstrap. CRC Press.

Finni, T., Uusi-Vähälä, M., Pesola, A.J., and Taipale, R.S. 2016. Do running and strength exercises reduce daily muscle inactivity time? AIMS Public Health 3: 702.

Hall, D.B., and Bailey, R.L. 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. Forest Science 47: 311–321.

Haskell, W.L., Yee, M.C., Evans, A., and Irby, P.J. 1993. Simultaneous measurement of heart rate and body motion to quantitate physical activity. Medicine and Science in Sports and Exercise 25: 109–115.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Hiilloskorpi, H., Pasanen, M., Fogelholm, M., Laukkanen, R.M., and Mänttäri, A. 2003. Use of heart rate to predict energy expenditure from low to high activity levels. International Journal of Sports Medicine 24: 332–336.

Kesaniemi, Y.A., Danforth, E., Jensen, M.D., Kopelman, P.G., Lefèbvre, P., and Reeder, B.A. 2001. Dose-response issues concerning physical activity and health: an evidence-based symposium. Medicine & Science in Sports & Exercise 33: S351–S358.

Keytel, L., Goedecke, J., Noakes, T., Hiiloskorpi, H., Laukkanen, R., van der Merwe, L., and Lambert, E. 2005. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. Journal of Sports Sciences 23: 289–297.

Laird, N.M., and Ware, J.H. 1982. Random-effects models for longitudinal data. Biometrics 38: 963–974.

Lappi, J. 1991. Calibration of height and volume equations with random parameters. Forest Science 37: 781–801.

Levine, J.A. 2005. Measurement of energy expenditure. Public Health Nutrition 8: 1123–1132.

Li, R., Deurenberg, P., and Hautvast, J. 1993. A critical evaluation of heart rate monitoring to assess energy expenditure in individuals. The American Journal of Clinical Nutrition 58: 602–607.

Lindstrom, M.J., and Bates, D.M. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46: 673–687.

Livingstone, M. 1997. Heart-rate monitoring: the answer for assessing energy expenditure and physical activity in population studies? British Journal of Nutrition 78: 869–871.

Livingstone, M., Prentice, A.M., Coward, W., Ceesay, S.M., Strain, J.J., McKenna, P.G., Nevin, G.B., Barker, M.E., and Hickey, R. 1990. Simultaneous measurement of free-living energy expenditure by the doubly labeled water method and heart-rate monitoring. The American Journal of Clinical Nutrition 52: 59–65.

Livingstone, M., Robson, P.J., and Totton, M. 2000. Energy expenditure by heart rate in children: an evaluation of calibration techniques. Medicine and Science in Sports and Exercise 32: 1513–1519.

Luke, A., Maki, K.C., Barkey, N., Cooper, R., and McGee, D. 1997. Simultaneous monitoring of heart rate and motion to assess energy expenditure. Medicine and Science in Sports and Exercise 29: 144–148.

Lusk, G. 1924. Animal calorimetry: twenty-fourth paper. Analysis of the oxidation of mixtures of carbohydrate and fat. Journal of Biological Chemistry 59: 41–42.

McArdle, W.D., Katch, F.I., and Katch, V.L. 2015. Exercise Physiology: Nutrition, Energy, and Human Performance. Lippincott Williams & Wilkins, Baltimore, 8th edition.

Mehtätalo, L., and Lappi, J. 2020. Biometry for Forestry and Environmental Data: with Examples in R. Chapman & Hall / CRC Press, Boca Raton, 1st edition, 411 p.

Meng, S.X., and Huang, S. 2009. Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. Forest Science 55: 238–248.

Moon, J.K., and Butte, N.F. 1996. Combined heart rate and activity improve estimates of oxygen consumption and carbon dioxide production rates. Journal of Applied Physiology 81: 1754–1761.

Oja, P., Ilmarinen, J., and Louhevaara, V. 1982. Heart rate as an estimator of oxygen consumption during manual postal delivery. Scandinavian Journal of Work, Environment & Health 8: 29–36.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team 2020. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-147.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-effects Models in S and S-plus. Statistics and Computing, Springer, New York, NY.

R Core Team 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rennie, K.L., Hennings, S.J., Mitchell, J., and Wareham, N.J. 2001. Estimating energy expenditure by heart-rate monitoring without individual calibration. Medicine and Science in Sports and Exercise 33: 939–945.

Schrack, J.A., Zipunnikov, V., Goldsmith, J., Bandeen-Roche, K., Crainiceanu, C.M., and Ferrucci, L. 2014. Estimating energy expenditure from heart rate in older adults: a case for calibration. PLoS ONE 9: e93520.

Schulz, S., Westerterp, K., and Brück, K. 1989. Comparison of energy expenditure by the doubly labeled water technique with energy intake, heart rate, and activity recording in man. The American Journal of Clinical Nutrition 49: 1146–1154.

Spurr, G., Prentice, A., Murgatroyd, P., Goldberg, G., Reina, J., and Christman, N. 1988. Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry. The American Journal of Clinical Nutrition 48: 552–559.

Tikkanen, O., Haakana, P., Pesola, A.J., Häkkinen, K., Rantalainen, T., Havu, M., Pullinen, T., and Finni, T. 2013. Muscle activity and inactivity periods during normal daily life. PLoS ONE 8: e52228.

Tikkanen, O., Kärkkäinen, S., Haakana, P., Kallinen, M., Pullinen, T., and Finni, T. 2014. EMG, heart rate, and accelerometer as estimators of energy expenditure in locomotion. Medicine and Science in Sports and Exercise 46: 1831–1839.

Tikkanen, O., Sipilä, S., Kuula, A.S., Pesola, A., Haakana, P., and Finni, T. 2016. Muscle activity during daily life in the older people. Aging Clinical and Experimental Research 28: 713–720.

Appendix A: Calibration of the logistic mixed model

The description of the calibration process of the logistic mixed model for an individual i who has not been a part of the original data (Hall and Bailey, 2001; Meng and Huang, 2009). After fitting the logistic model for the original data, estimates of the fixed effects and parameters related to the covariance matrices are marked by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, respectively. The prediction of $\hat{\mathbf{u}}_i$ is performed by an algorithm where random effect values are updated iteratively. Denote the iteration number by ω and the derivative of the logistic function $g(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{ij})$ with respect to $\mathbf{u}_i$ at iteration ω by

$$\hat{\mathbf{Z}}_i^{(\omega)} = \begin{bmatrix} \frac{\partial g}{\partial u_1}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{i1}) & \frac{\partial g}{\partial u_2}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{i1}) & \frac{\partial g}{\partial u_3}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{i1}) \\ \vdots & \vdots & \vdots \\ \frac{\partial g}{\partial u_1}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{is}) & \frac{\partial g}{\partial u_2}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{is}) & \frac{\partial g}{\partial u_3}(\mathbf{A}_i\hat{\boldsymbol{\beta}}+\mathbf{u}_i, HR_{is}) \end{bmatrix},$$

where the number of rows in a matrix equals the number of observations (denoted by s) used in calibration.

In the following algorithm $\mathbf{EE}_i$ stands for the observed energy expenditures at chosen calibration loads. The iterative algorithm consists of four steps:

1. Let $\omega = 0$. Define $\mathbf{u}_i^0 = \mathbf{0}$ and calculate $\hat{\mathbf{Z}}_i^{(0)}$.

2. Let $\omega = 1$. Predict random effects with equation

$$\mathbf{u}_i^1 = \hat{\mathbf{D}}\hat{\mathbf{Z}}_i^{(0)\intercal}(\hat{\mathbf{R}}_i + \hat{\mathbf{Z}}_i^{(0)}\hat{\mathbf{D}}\hat{\mathbf{Z}}_i^{(0)\intercal})^{-1}(\mathbf{EE}_i - \boldsymbol{g}(\mathbf{A}_i\hat{\boldsymbol{\beta}}, \mathbf{HR}_i)).$$

3. Calculate $\hat{\mathbf{Z}}_i^{(\omega)}$ and $\boldsymbol{g}(\mathbf{A}_i\hat{\boldsymbol{\beta}} + \mathbf{u}_i^{(\omega)})$. Update random effects with equation

$$\mathbf{u}_i^{(\omega+1)} = \hat{\mathbf{D}}\hat{\mathbf{Z}}_i^{(\omega)\mathsf{T}}(\hat{\mathbf{R}}_i + \hat{\mathbf{Z}}_i^{(\omega)}\hat{\mathbf{D}}\hat{\mathbf{Z}}_i^{(\omega)\mathsf{T}})^{-1}(\mathbf{EE}_i - \boldsymbol{g}(\mathbf{A}_i\hat{\boldsymbol{\beta}} + \mathbf{u}_i^{(\omega)}, \mathbf{HR}_i) + \hat{\mathbf{Z}}_i^{(\omega)}\mathbf{u}_i^{(\omega)}).$$

4. Repeat step 3 until a chosen convergence criterion is fulfilled:
   for instance max($|u_{i1}^{(\omega+1)} - \hat{u}_{i1}^{(\omega)}|, |u_{i2}^{(\omega+1)} - u_{i2}^{(\omega)}|, |u_{i3}^{(\omega+1)} - u_{i3}^{(\omega)}|) < 0.00001$. Let us denote the last value by $\hat{\mathbf{u}}_i$.

EE can now be predicted at HR value x using function $g(\mathbf{A}_i\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_i, x)$.

Appendix B: A parametric bootstrap algorithm

An algorithm for computing confidence and prediction intervals for logistic and linear mixed models, by using a general parametric bootstrap algorithm (Efron and Tibshirani, 1994). In the following let us denote the logistic function by g, data without individual i by $\mathbf{y}^{(-i)}$ and other symbols similarly, and individual i's observations used in calibration by $\mathbf{y}_i^{(cal)}$. The prediction and confidence intervals of EE for individual i were computed using the following algorithm:

1. Estimate parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on data $\mathbf{y}^{(-i)}$ and mark the estimates with symbols $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$.

2. Sample randomly $\boldsymbol{\epsilon}^{*(-i)} \sim N(\mathbf{0}, \mathbf{R}(\hat{\boldsymbol{\theta}}))$ and $\mathbf{u}^{*(-i)} \sim N(\mathbf{0}, \mathbf{D}(\hat{\boldsymbol{\theta}}))$.

3. Form new data $\mathbf{y}^{*(-i)} = \boldsymbol{g}(\mathbf{A}^{(-i)}\hat{\boldsymbol{\beta}} + \mathbf{u}^{*(-i)}, \mathbf{HR}^{(-i)}) + \boldsymbol{\epsilon}^{*(-i)}$ in the logistic case or $\mathbf{y}^{*(-i)} = \mathbf{X}^{(-i)}\hat{\boldsymbol{\beta}} + \mathbf{Z}^{(-i)}\mathbf{u}^{*(-i)} + \boldsymbol{\epsilon}^{*(-i)}$ in the linear case.

4. Fit the model for data $\mathbf{y}^{*(-i)}$ and mark the stored model parameters as $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\theta}}^*$.

5. Predict $\mathbf{u}_i^{**}$ at individual level for individual i using the calibration measurements $\mathbf{y}_i^{(cal)}$ and sample randomly $\mathbf{u}_i^{**} \sim N(\mathbf{0}, \mathbf{D}(\hat{\boldsymbol{\theta}}^*))$ at population-level, instead.

6. Predict EE at individual and population-level for an individual i at required HR values $\mathbf{x}_i$ using $\boldsymbol{g}(\mathbf{A}_i\hat{\boldsymbol{\beta}}^* + \mathbf{u}_i^{**}, \mathbf{x}_i)$ or $\mathbf{X}_i\hat{\boldsymbol{\beta}}^* + \mathbf{Z}_i\mathbf{u}_i^{**}$.

7. Add simulated error term $\boldsymbol{\epsilon}_i^* \sim N(\mathbf{0}, \mathbf{R}_i(\hat{\boldsymbol{\theta}}^*))$ to the predictions.

8. Repeat steps 2 - 7 K times.

9. Report the mean prediction given HR at individual and population-level.

10. Report the confidence and prediction intervals given HR at individual and population-level taking, for example, the 5th and 95 the percentiles of the realisations.