THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Modelling a response as a function of high frequency count data: the association between physical activity and fat mass

OPEN ACCESS

Modelling a response as a function of high-frequency count data: the association between physical activity and fat mass

Nicole H. Augustin,[*,1] Calum Mattocks,[2] Julian J. Faraway,[1]
Sonja Greven[3] and Andy R. Ness[4]

[1] Department of Mathematical Sciences, University of Bath, Bath, UK;
[2]Centre for Exercise, Nutrition and Health Sciences,
School for Policy Studies, 8 Priory Road,
University of Bristol, Bristol, UK;
[3]Department of Statistics, Ludwig-Maximilians-University Munich,
Ludwigstr. 33, 80539 Munich, Germany;
[4]School of Oral and Dental Science and School of Social and Community
Medicine, Bristol Dental School, Lower Maudlin Street, Bristol, UK

July 16, 2015

Abstract

Accelerometers are widely used in health sciences, ecology and other application areas. They quantify the intensity of physical activity (PA) as counts per epoch over a given period of time. Currently, health scientists use very lossy summaries of the accelerometer time series, some of which are based on coarse discretisation of activity levels, and make certain implicit assumptions, including linear or constant effects of PA. We propose the histogram as a functional summary for achieving a near lossless dimension reduction, comparability between individual time series and easy interpretability. Using the histogram as a functional summary avoids registration of accelerometer counts in time. In our novel method a scalar response is regressed on additive multi-dimensional functional predictors, including the histogram of the high-frequency counts, and additive non-linear predictors for other continuous covariates. The method improves on the current state-of-the art, as it can deal with high-frequency time series of different lengths and missing values and yields a flexible way to model the PA effect with fewer assumptions. It also allows the commonly made modelling assumptions to be tested. We investigate the relationship between the response fat mass and physical activity measured by

*Nicole H. Augustin, n.h.augustin@bath.ac.uk

accelerometer, in data from the Avon Longitudinal Study of Parents and Children (ALSPAC). Our method allows testing of whether the effect of PA varies over its intensity by gender, by time of day or by day of the week. We show that meaningful interpretation requires careful treatment of identifiability constraints in the light of the sum-to-one property of a histogram. We find that the (not necessarily causal) effect of PA on kg fat mass is not linear and not constant over the activity intensity.

1 Introduction

This paper proposes a novel method for practical and interpretable modelling of predictors that are high-frequency monitoring of counts or intensity that contain much more information on the data than the current methods. Examples of high-frequency count data include the monitoring of traffic, heart rate in intensive care, brain activity and financial transactions. We are concerned with data on the intensity of physical activity (PA) recorded by an accelerometer. Accelerometers are now widely used for many different applications, such as protecting hard drives from impact damage, monitoring vibrations in industry and in many medical situations. The specific accelerometer used for the application discussed here records the time-varying acceleration signal. The signal passes through a filter that passes frequencies within a range typical for human motion (10 to 50 Hz) and excludes frequencies atypical for human motion such as vibrations from a car. Due to limited storage, the signal - measured in units called counts - is summed over a user-specified interval called an epoch. The resulting data are hence counts per epoch, ranging between 10 - 60 seconds, over a given period of time, e.g. a number of weeks. Figure 1 shows four sample accelerometer profiles of children aged 12 years with counts summarised by minutes. Due to data storage becoming smaller and cheaper, newer generation accelerometers are able to store the pass band filtered raw signal without summing it over epoch; this leads to even denser time series of PA measurements.

In epidemiology accelerometers are now increasingly used to monitor the intensity of PA for investigating its relationship with adiposity, cardiovascular disease, diabetes, depression and other health outcomes. For estimating energy expenditure from accelerometer output regression methods are traditionally used. More recently, machine-learning algorithms - such as support vector machines [9, 43] , artificial neural networks [32, 41] and hidden Markov models [24] - have been used for estimating energy expenditure and classifying type of activity. All these methods rely on initial calibration studies that record energy expenditure and accelerometer counts for different type of activities. Mostly scalar summaries of accelerometer data are used as predictors. Other work applies functional data analysis methods to smoothed accelerometer data as functions of time as a response, for example Morris et al. [22] used wavelet-based functional mixed models, Sera et al. [30] used functional analysis of variance models based on penalised splines; see also [4, 34] and [7, 40, 31] for various approaches to functional principal component analysis. As daily structure is highly variable between individuals it is not clear whether and how the accelerometer time series

are registered in time in order to make the activity curves comparable between individuals. The above approaches are useful for exploring variability in the accelerometer time series by time of day, season, environmental and socio-economic factors, but it is not clear how the methods could be used to describe the effects of a PA time series on a response in an interpretable way.

In the epidemiological setting where PA may be the predictor of a health outcome, the high-dimensional accelerometer time series are typically summarised into a single summary statistic per individual. Examples for such statistics are: mean counts per minute (cpm), mean daily moderate to vigorous PA (MVPA) (which is the average minutes per day spent at moderate or vigorous activity), and average sedentary behaviour (which is the average minutes per day spent in sedentary activity) [28, 21, 20]. For each of these summaries, cut-points of counts per minute are used for different levels of energy expenditure relating to light, moderate and vigorous activity. These are based on estimates of energy expenditure described above, either using published models from the literature or using models based on calibration studies performed on a subset of the study population. There is some debate whether the energy expenditure and the cut-points change with age in children [27]. Using only scalar summaries ignores the pattern of PA, that is the distribution of the activity counts per minute and the frequency spectrum of activity counts.

Figure 1 illustrates that accelerometer profiles typically are high-frequency spiky time series containing missing periods, here ranging from minutes to whole days. In order to compare accelerometer profiles between individuals and to reduce the high-dimension a functional summary is required. When mean cpm and MVPA are used as predictors in models certain implicit assumptions about the effect of PA on the outcome are made. We propose to use instead the histogram as a functional summary. The histogram automatically deals with high-frequency time series of differing lengths and with missing values, but still contains much of the information in the data. This yields a flexible way to model the PA effect on the outcome with fewer assumptions. It also allows the commonly made modelling assumptions to be tested. We can easily extend the model to allow for interactions between PA and gender, day of the week, or time of day and other variables. The model also allows the inclusion of smooth effects of scalar predictors.

In our novel method a scalar response is regressed on additive multi-dimensional functional predictors for exploring the relationship between high-frequency count data of PA and health outcomes. The method is an extension of generalised regression of scalars on functions or signal regression [14, 26] and extends the approach presented in [2]. We use the modelling approach to investigate the following questions:

1. Is the effect of PA linear over the activity intensity range, i.e. is it sufficient to use mean cpm as a predictor? Or does the effect of PA vary over the intensity range and how?

2. Does the effect of PA intensity vary by gender? This would be plausible as there are differences in the metabolism.

3. Does the effect of PA intensity vary by time of day?

4. Does the effect of PA intensity vary depending on whether it happens on a weekday or on the weekend? Evidence for different effects could be explained by the fact that weekends and weekdays are associated with different activity types.

We investigate the above questions using data from ALSPAC where accelerometer data is available as a time series of accelerometer counts per minute over seven days for a subset of children. Our interest lies in fat mass as a health outcome. At ages of approximately 12, 14 and 16 years the children were attending research clinics where they were asked to wear an accelerometer for seven days. Fat mass was also assessed using a DXA scanner. We use fat mass as a health outcome because it is a measure of adiposity and it is strongly associated with other health outcomes such as cardio-metabolic health (cardiovascular disease and type 2 diabetes) [11]. We consider fat mass at age 12, 14 and 16 years with physical activity at 12, 14 and 16 years as predictors, since lack of physical activity in youth is a likely cause of early onset adiposity which is associated with adiposity in adults [42]. We use fat mass rather than body mass index (BMI) because it is more precise - BMI does not differentiate between fat and lean mass.

Regarding the association between fat mass and PA, the ALSPAC data have been extensively analysed by considering summary statistics such as MVPA and mean cpm as predictors for the health outcome fat mass and vice versa. The results for ages 12 and 14 years indicate that an increase in MVPA is associated with a lower risk of obesity in children as they go through adolescence and more so in boys than girls [28]. When the analyses were repeated with BMI as outcome, similar patterns of association were found, but associations estimated from models with the outcome BMI were considerably weaker than associations estimated from models with the outcome fat mass [23, 28]. As in Riddoch et al. [28] and due to the unknown lag structure and interaction with age of potential effects of PA on fat mass, we fit several cross-sectional and prospective models to the ALSPAC data resulting in six different age combinations of fat mass and PA. This allows us to assess whether the effects of PA differ between ages and whether there are any prospective associations between PA and fat mass, which would be somewhat indicative of causality. We also use the model to predict changes in fat mass when we redistribute the time spent in different intensity levels of PA. These predictions also show how our model can be used to investigate potential changes regarding PA.

In section 2, we introduce the data and the processing protocol and present summary statistics. In section 3, we present the regression of a scalar on multi-dimensional functional and additive non-linear predictors; we cover parameter estimation, model selection and extra considerations for model interpretation induced by the sum-to-one property of the histogram. In section 4 we present results followed by a discussion in section 5.

2   The ALSPAC data

ALSPAC is a birth cohort study, see Boyd et al. [3] and Fraser et al. [6] for a detailed description. In this study, all pregnant women in the former Avon Health Area who had an expected delivery date between 1/04/1991 and 31/12/1992, were asked to participate in the study. The Avon Health Area is situated around Bristol in the UK. A total of 14,541 pregnant women were enrolled, and this resulted in 14,062 live births. Detailed data have been collected from pregnancy onward. Please note that the study website contains details on data that is available at `www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/`. All chil-

dren of the study have been invited to regular research clinics from the age of 7 years. At the ages of 11, 13 and 15 years the children attending these clinics were asked to wear an MTI Actigraph AM7164 2.2 Accelerometer for seven days. As the mean ages of children were actually closer to 12, 14 and 16 years we will refer to the children as 12-, 14- and 16-year-olds. Children wore the accelerometer at the waist during waking hours, except for showering, bathing, swimming and any watersports. At age 14, children were wearing either the AM7164 or the newer GT1M Actigraph. Unfortunately data on which type of accelerometer was worn by the children at age 14 was not available for this analysis. In addition fat mass was derived using a Lunar Prodigy DXA scanner (GE Medical Systems Lunar, Madison, WI, USA) and the children's height was measured. See Mattocks et al. [18] for more details on the measurement protocol.

2.1   Summary statistics

Here we consider fat mass and activity profiles at all three ages of 12, 14 and 16 years. Not all of the children who had their fat mass measured by scanner also had their PA measured and vice versa. Following the processing protocol of the accelerometer data yields 4161 children with both measurements at age 12, 2919 children at age 14 and 1488 children at age 16. Of the 4161 children with both measurements at age 12, 2484 had both measurements repeated at age 14 and 1352 at age 16. Below we use the accelerometer data in the processed form, that is blocks of 10 zeros and invalid days have been set to missing. For details see the Supplementary Material.

The distribution of the outcome variable fat mass (kg) is right-skewed (Figure 2). Girls have a higher fat mass than boys and this difference becomes more pronounced as they go through adolescence. Figure 3 shows the mean histogram of PA over all individuals and ages. A plot of the difference between the mean histogram of each age by gender combination (not shown) gives the following insight: at the same age boys, on average, are more active than girls. The biggest differences in intensity distribution are seen between ages: for example, at age 12, children spent the smallest amount of time in sedentary activity (very low to zero cpm) and spent more time in moderate to vigorous activity. With increasing age boys and girls are on average less active. Investigating the mean of the two dimensional histograms over PA intensity and hour of the day shows, that time spent in sedentary activity increases after 14:00h for all all ages. Figure 4 shows the correlation between the relative frequency of the histogram bins for individuals at age 12, 14 and 16. Shown is the correlation of the columns of a matrix containing the individual histograms by row, where the rows represent individuals and the columns the histogram bins. Besides very strong positive correlation between density values of neighbouring bins, there is negative correlation up to -0.6 between sedentary behaviour (below 200 cpm) and light activity (above 200 cpm). For all three ages the correlation between sedentary and moderate to vigorous activity is very small. This implies that if individuals spent little time in sedentary activity they spent more time in light activity and vice versa.

3 Models

We introduce the approach by considering the following generalised linear model (GLM) for fat mass as a function of the accelerometer profile

$$g(\mu_i) = x_i^T \boldsymbol{\beta} + \sum_j \gamma_j z_i(p_j).$$

The y_i is the scalar response with mean μ_i for individual i - (log) total fat mass at a given age for the ALSPAC study - $z_i(p_j)$ is the relative frequency of the histogram with some given number at mid-point p_j and derived from the vector of count entries. The g is a known monotonic link function. The y_i follow an exponential family distribution and the vector x_i is a row of the model matrix containing scalar variables and factors - in our case `intercept`, `gender`, `height`, `height`^{2}, `weartime` and whether the mother is obese (`m.obese`). The variable `weartime` is included because it is a measure of compliance which could be a proxy for hidden confounders. The problem with this model is that, besides being unidentifiable due to $\sum z_i(p_j) = 1$, the coefficients γ_j for the histogram values will be highly correlated and their interpretation will hence be difficult. The model treats the relative frequency for each of the histogram mid-points as a separate explanatory variable and this is not appropriate. It makes more sense to exploit the fact that the histogram is a function and constrain the estimate of γ_j to be similar to the γ_l at neighbouring mid-points p_l. That is, the histogram enters the model as a function. We let the γ_j vary smoothly, where $\gamma_j = f(p_j)$, an unknown 'coefficient' function to be estimated and $f(p)$ can be parameterised using any penalised regression spline basis. Then

$$g(\mu_i) = x_i^T \boldsymbol{\beta} + \sum_j f_p(p_j) z_i(p_j).$$

Here $f_p(p)$ is represented using an adaptive smooth [36] with a P-spline basis [5], based on a B-spline basis function and discrete random walk penalties on the basis coefficients. The model is identifiable through a constraint on the smooth function $f_p(p)$, see our discussion in Section 3.3. The smooth is adaptive by letting the terms in the penalty have different weights depending on the histogram mid-point j and then multiplying the smoothness parameters by these weights. We use an adaptive smooth in order to allow different degrees of smoothness at different intensity levels. Models such as the above are also called generalised regression of scalars on functions or signal regressions. Marx and Eilers [14] first applied P-splines in signal regression; see also Ramsay and Silverman [26].

We extend the above model by also allowing multi-dimensional functional predictors and additive non-linear predictors for continuous covariates. This is illustrated by the following models for the ALSPAC study, each of which will be used to answer one of the four questions posed in the introduction. The first model below allows us to test whether the effect of intensity is linear over the activity range.

$$\text{Baseline model} \qquad g(\mu_i) = x_i^T \boldsymbol{\beta} + f_h(\texttt{height}_i) + \sum_j f_p(p_j) z_i(p_j) \qquad (1)$$

where f_h is a smooth function and the vector x_i contains `intercept`, `gender`, `weartime` and `m.obese`. Note that the way we specify the effect of PA actually allows us to check whether the

common modelling approach using the scalar summary statistic mean cpm is adequate. In order to see this, we point out that model (1) contains as a special case for a linear effect function $f_p()$ the model

Question 1
$$g(\mu_i) = x_i^T \boldsymbol{\beta} + f_h(\texttt{height}_i) + \gamma \sum_j p_j z_i(p_j). \qquad (2)$$

As the mean cpm can be written as $\sum_j p_j z_i(p_j)$ estimated from the histogram, the models are nested and allow, for example, for testing of the implicit assumptions made in using mean cpm as a predictor. Further we can investigate whether the effect of PA intensity varies by gender by introducing different smooth functions for males and females:

Question 2
$$g(\mu_i) = x_i^T \boldsymbol{\beta} + f_h(\texttt{height}_i) + \sum_j f_{s(i)}(p_j) z_i(p_j) \qquad (3)$$

where $s(i) = 1$ for male and $s(i) = 2$ for female. For investigating whether the effect of PA intensity varies with time of day we introduce, $z_i(p_j, t_m)$, a two dimensional histogram over PA intensity with bins of equivalent width, as in model (1), and hour of the day with bin widths l_m:

Question 3
$$g(\mu_i) = x_i^T \boldsymbol{\beta} + f_h(\texttt{height}_i) + \sum_j \sum_m f_{pt}(p_j, t_m) z_i(p_j, t_m) l_m \qquad (4)$$

where the smooth function $f_{pt}(p_j, t_m)$ is now two-dimensional. And lastly, we investigate whether the effect of PA intensity varies between weekday and weekend.

Question 4 $$g(\mu_i) = x_i^T \boldsymbol{\beta} + f_h(\texttt{height}_i) + \sum_j f_{we}(p_j) z_i^{we}(p_j) + \sum_j f_{wd}(p_j) z_i^{wd}(p_j) \qquad (5)$$

where f_{we} is the smooth function over the histogram for the weekend and f_{wd} for weekdays, with separate histograms for weekend $z_i^{we}(p_j)$ and weekday $z_i^{wd}(p_j)$.

3.1   Parameter estimation

Parameter estimation of the above regression with additive multi-dimensional functional predictors is based on the fact that all of the models 1 to 5 can be expressed as a penalised GLM; $g(\mu_i) = \mathbf{X}_i \boldsymbol{\theta}$, where $\mathbf{X}_i$ is a row of the model matrix containing all components of the model; that is, all strictly parametric components, such as explanatory variables of linear effects, the sum of the relative frequencies of the histogram of observation i multiplied by the basis functions evaluated at histogram mid-points p_j and the basis functions for non-linear predictors evaluated at observations i. The parameter vector $\boldsymbol{\theta}$ contains the coefficients of the bases functions and coefficients of linear terms. Each of the smooth functions used in models (1) to (5) has one or several quadratic penalties, $\boldsymbol{\theta}^T \boldsymbol{S}^k \boldsymbol{\theta}$, with an associated smoothing parameter λ_k and the matrices $\mathbf{S}^k$, contain known coefficients of the penalty. Typical penalty choices are derivative-based and the penalty thus measures how wiggly the smooth function is and is defined by the type of smooth. The model is then estimated by minimising $D(\boldsymbol{\theta}) + \sum_{k=1}^P \lambda_k \boldsymbol{\theta}^T \boldsymbol{S}^k \boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$, where D is the model deviance:

the saturated log likelihood minus the log likelihood all multiplied by two times the scale parameter ϕ. In models 1 to 5 we made the following choices for smooths. The smooth f_h for height is represented using a thin plate regression spline [38] with a penalty based on the second derivative of the smooth function. The adaptive smooth used for the histogram (f_p, $f_{s(i)}$, f_{we} and f_{wd}) has adaptive penalties along the intensity of PA. The two-dimensional smooth function $f_{pt}(p_j, t_m)$ in model (4), expressed by a tensor product of two cubic regression spline bases, has separate smoothing parameters and second derivative penalties for the two dimensions intensity and time. By having two penalties and smoothing parameters we ensure that the smoothness is appropriate for the two scales. Tensor product smooths are useful for multi-dimensional smoothing problems, where the dimensions represent different scales [1].

The parameters are estimated using a nested iteration scheme [36] where the outer iteration is approximate restricted maximum likelihood (REML) estimation of smoothness parameters, minimising a Laplace approximation of REML where the parameter vector, $\boldsymbol{\theta}$, containing coefficients of linear terms and of basis functions for the smooth terms, is integrated out, see e.g. Laird and Ware [12]. The inner iteration is a penalised iterative re-weighted least squares (PIRLS) algorithm to find all other parameters, i.e. the coefficients of basis functions, and coefficients of linear terms. We use the computational set up of Wood [36] implemented in the `gam()` function of the `R` package `mgcv` [25] to estimate the model.

For prediction we use Bayesian credible intervals by sampling from the posterior of all model parameters $\boldsymbol{\theta}$ to obtain a sample from the predictive distribution, see also Wahba [33], Wood [38]. For this, the model is represented as a Bayesian model, where the smooth terms are a mixture of fixed and random effects. This approach recognises that imposing a particular penalty effectively imposes some prior beliefs about the likely characteristics of the correct model. That is, the model structure allows considerably more flexibility than is believed to be really likely, and the choice is made to penalise models that are in some sense too wiggly. This is done by specifying a prior distribution on the parameters $\boldsymbol{\theta}$ with a distribution proportional to $\exp(-\frac{1}{2}\boldsymbol{\theta}^T \sum_{k=1}^{P} \mathbf{S}^k \tau_k \boldsymbol{\theta})$, i.e. the quadratic penalties are equivalent to assuming an improper multivariate normal prior. The τ_k are parameters controlling the dispersion of the prior, and the matrices $\mathbf{S}_k$ contain known coefficients of the P penalties. Choosing the τ_k in this Bayesian formulation is equivalent to choosing the smoothing parameters λ_k since $\tau_k = \frac{\lambda_k}{\phi}$, where ϕ is the scale parameter. Here the choice of the λ_k is data driven as they are estimated by REML.

With the generally improper prior distribution for $\boldsymbol{\theta}$, a posterior distribution for $\boldsymbol{\theta}$ can be derived: $\boldsymbol{\theta} \sim \mathrm{N}(\hat{\boldsymbol{\theta}}, (\mathbf{XWX}^{\mathsf{T}} + \sum_k \lambda_k \mathbf{S}^k)^{-1}\phi)$, with $\mathbf{W} = \mathrm{diag}\frac{1}{V(\boldsymbol{\mu})g(\boldsymbol{\mu}^2)}$, see Wood [38] for details. We obtain a posterior sample of the distribution of estimates $\boldsymbol{\theta}$ and obtain from this the predictive distribution of the response. Then the pth draw from the posterior distribution is $\mu_{ip} = \hat{y}_{itp} = g^{-1}\left(\mathbf{X}_i \hat{\boldsymbol{\theta}}_p\right)$. For prediction the $\mathbf{X}_i$ is re-evaluated for the prediction data. Otherwise the same procedure as described above is carried out. E.g. for estimating the percentage change in fat mass when some of the sedentary activity is redistributed to higher intensities, as shown in Figure 6, we obtain a sample of the percentage change $r_{ip} = \left(\frac{g^{-1}(\mathbf{X}_i^r \hat{\boldsymbol{\theta}}_p)}{g^{-1}(\mathbf{X}_i \hat{\boldsymbol{\theta}}_p)} - 1\right) \times 100$ where $\mathbf{X}_i$ is the observed model matrix for individual i and $\mathbf{X}_i^r$ is the model matrix where relative frequencies of the histogram are redistributed according to some scenario. Shown in Figure 6 are the median and the 2.5% and 97.5% percentiles of the posterior distribution of r_{ip}.

8

3.2 Model selection and validation

We use the root mean square prediction error (RMSPE) estimated by validation as a selection criterion. In addition we consider the Akaike information criterion (AIC), Bayesian information criterion (BIC) and adjusted R^2. The AIC is defined as $AIC = -2l(\hat{\boldsymbol{\theta}}) + 2edf$, the BIC is $BIC = -2l(\hat{\boldsymbol{\theta}}) + \log(n)edf$, where l is the conditional log likelihood given the penalised parameters, edf are the effective degrees of freedom, here estimated by the trace of the matrix that maps the unpenalised estimates onto the penalised estimates corrected for the smoothing parameter uncertainty as described in Wood et al. [39]; see also Greven and Kneib [8]. Comparing these criteria between model (1) and model (2) will show how much can be gained by allowing the effect of PA intensity to be non-linear. We also test formally whether the effect for PA is non-linear using a significance test. Replacing the adaptive smooth basis with a thin plate regression spline basis makes this task relatively easy. The linear part of the regression spline basis, that is the completely smooth basis function with zero penalty, is separate from the rough penalised basis functions. Hence the linear basis part can be included as a separate effect in a model that also includes the rough penalised basis function. In this nested model we test whether the non-linear part of the basis is zero using an approximate likelihood ratio test [37].

3.3 Model interpretation

Due to the histogram property, we have that $\sum_j z_i(p_j) = 1$, yielding equal model fits for $\alpha + \sum_j f(p_j)z_i(p_j)$ and $(\alpha + c) + \sum_j (f(p_j) - c)z_i(p_j)$ for any c. We thus need an additional identifiability constraint for our model. There are several ways this could be achieved, with differing corresponding model interpretations. The standard assumption is a sum-to-zero constraint $\sum_j f(p_j) = 0$, and this is the default as implemented in the in the `gam()` function of the R ([25]) package `mgcv` [36]. To help interpret the coefficient function $f(p_j)$ under this constraint in model (1) and its variant models 3 to 5, consider the simpler model $\alpha + \sum_j f(p_j)z(p_j)$. In this model, due to $\sum_j f(p_j) = 0$, the intercept α can be interpreted as the expected fat mass on the link function scale $g(\mu_i)$ for the scenario that PA is equally distributed along intensity with $z_i(p_j) = c$ for some c inversely related to the number of bins. The functional coefficient $f(p_j)$ measures the deviation from this expected value according to the actual PA distribution.

A second possible constraint is to restrict $f(p_1) = 0$ for the first bin from the histogram corresponding to sedentary activity. If we insert $z_i(p_1) = 1 - \sum_{j>1} z_i(p_j)$ into the model above, then $\alpha + f(p_1)z_i(p_1) + \sum_{j>1} f(p_j)z_i(p_j) = \alpha + f(p_1) + \sum_{j>1} (f(p_j) - f(p_1)) z_i(p_j) = \alpha' + \sum_{j>1} f'(p_j)z_i(p_j)$. This shows that if we set $f(p_1) = 0$, or equivalently remove the first histogram bin in our model fit, the intercept is now $\alpha' = \alpha + f(p_1)$, which is the expected value of fat mass on the link function scale for sedentary activity, as defined by the breaks of the first histogram bin. The interpretation of the functional coefficient is also more convenient: $f'(p_j) = (f(p_j) - f(p_1))$ is the effect of activity compared to sedentary activity. Hence in the following analysis we fit all models with the first histogram bin removed instead of imposing the centred smooth constraint to achieve this convenient interpretability.

4 Results and interpretation

We fit models (1) to (5) for the six different age combinations of the variables to investigate possible age-varying effects with unknown lag-structure of PA on fat mass. We model the log of fat mass using a normal distribution as this gives a better fit compared to modelling fat mass using a gamma distribution with a log link function. We compare these models with the different criteria in Table 1 to model 'base' ($\log(y_i) = x_i^T \boldsymbol{\beta} + f_h(height_i) + \epsilon_i$) which does not include a term for PA. We use the following short notations (a) 12-12, (b) 12-14, (c) 12-16, (d) 14-14, (e) 14-16 and (f) 16-16, where the first number refers to the age when PA was measured and the corresponding weartime in the model and the second number refers to the age of the response fat mass and covariate height in the model. E.g. (b) 12-14 has PA and weartime at age 12 and response fat mass and covariate height at age 14. The models are fitted to approximately 3/4 of training data and the root mean squared prediction error (RMSPE) is then estimated on the remaining 1/4 of validation data. Note that the sample size N_T of the training data depends on whether both fat mass and PA was measured for the specific age combination. All the results are based on using 80 bins for the individual histograms of the accelerometer counts. We selected the number of bins which minimised the prediction error resulting in 80 bins of equal width, with the exception of the last bin which contains all counts in the range [8000,15000). Counts greater than 15000 are set to missing as misrecording is the most likely cause for these (see Supplementary Material). The results show that the effect of PA intensity is not linear since model (1) is always better than model (2) which only uses mean count per minute. The results show that for most age combinations model (1) is best, and there is little evidence that any of the more complex models (3) to (5) provides a better fit. In (b) 12-14 models (3) and (4) yield slightly lower RMSPEs than model (1) and in (f) 16-16, the AIC, BIC, and RMSPE select different best models, with model 5 yielding the lowest RMSPE. Hypothesis testing, as described in Section 3.2 on whether the effect of PA is non-linear confirms the non-linearity. The p-values for the test of whether the non-linear part of the basis is zero are all $< 3e-08$ except for model (f) 16-16, where the p-value was not significant.

These results show the assumption of a linear effect of PA over intensity made by using the summary mean (cpm) is not valid. Similarly, using MVPA as a predictor, as e.g. in Riddoch et al. [28], assumes zero effect of activity for an intensity below moderate and a constant effect of activity at and above the assumed cut-point for moderate intensity. Figure 5 shows the effect of the histogram of the accelerometer profile, a functional predictor. This clearly shows that the effect of PA is not linear and not constant over intensity. Our comparison of the RMSPE estimated by validation confirms that the proposed models yield a lower prediction error than model (2), a model which is often used in the literature. See also Augustin et al. [2] with similar findings based on a subset of the ALSPAC data. Studies have shown that there are differences in the output produced by different generations of Actigraph accelerometer see for example [29, 10]. At 12 and 16 years children were wearing the AM7164 and at age 14, children were wearing either the AM7164 or the newer GT1M Actigraph. Unfortunately data on which type of accelerometer was worn by the children at age 14 was not available for this analysis and hence we could not further investigate this. Interestingly, the effects of models involving PA measured at age 14, (d) 14-14 and (e) 14-16, are very similar to the effect estimated in the other models.

We can answer the four questions we posed in the introduction: There is evidence that the effect of

PA is not linear over the intensity range and that including the scalar summary statistic mean cpm in the model thus is not adequate to capture the true underlying relationship between PA and fat mass. Specifically there are significant associations between increased fat mass and PA at different lower light intensity levels (300 - 1000 cpm) as the 95% credible intervals of $f(z_j)$ did not include zero at these intensity levels. This intensity corresponds to light activities such as slow walking (e.g. to shops) or general hanging out with friends, which again may be associated with certain (eating) habits. In the prospective models (b, c, e) and the cross-sectional model at age 16 (f) we found that increased PA intensity of 2400 cpm and higher is associated with decreased fat mass (Figure 5). For the two cross-sectional models at age 12 and 14 (a, d), the confidence intervals are wider, but it appears that the decreasing effect on log fat mass and PA starts at a higher intensity of around 4000 cpm. It is important to note that these results do not imply causality as our data are observational. In fact causal effects in both directions between PA and fat mass are possible. In particular, adiposity may lead children to be less active.

For most age combinations there is little evidence to suggest that the effect of PA intensity varies by gender (model (3)), or whether it happens on weekdays or on weekends (model (5)) or by time of day (model (4)). For the age combinations (b) 12-14, the RMSPE of models (1), (3) and (4) are very close, and for the cross-sectional model at age 16 (f) model (5) has the lowest RMSPE. This is in contrast to Ness et al. [23] where an interaction between the effects of gender and MVPA was reported from an analysis of the ALSPAC data using a cross-sectional model with both obesity and PA at age 12. These different results can be explained by the fact that quite different models are fitted. In Ness et al. [23] the outcome variable is binary (obesity, yes/no). Here we use the log of fat mass as the outcome, since dichotomising fat mass into a binary obesity outcome and then using a Bernoulli distribution with a logit link, would lead to unnecessary loss of information. Our approach uses the histogram of PA counts as a predictor. Figure 3 showed that the mean histograms are quite different between boys and girls and our approach takes the shape of the PA intensity distribution into account. In Ness et al. [23] only MVPA is used as a predictor and the intensity distribution is not taken into account. The difference in estimates for MVPA in Ness et al. [23] for boys and girls may be due to the fact that the PA intensity distribution is different for boys and girls and that the distribution of intensity has an effect on obesity.

The resulting estimates of the coefficient function $f(p_j)$ for the best model (1) for each of the six different age combinations (a-f) in Figure 5 clearly show that the effect of PA is not linear or constant over intensity, as would be implied by the use of the common summary statistics mean cpm or MVPA. The $f(p_j)$ can be interpreted as a weight in the approximation to the integral of the histogram (with the first bin removed). With sedentary activity (counts 0 to 100 cpm) as a reference, the estimated weight function is positive in the range of the accelerometer profile, which has an increasing contribution to fat mass, and negative in the range of accelerometer counts with a decreasing contribution to fat mass. Compared to spending time in sedentary activity, time spent in activity with intensity above 1400 cpm has a decreasing effect on log fat mass. Between 200 and 800 cpm the effect of activity on log fat mass is increasing.

Using the 'best' model (1) we investigate how much variability in log fat mass the separate terms explain in Table 2. Overall, all models (1) have an adjusted R^2 between 28% and 36%. It is interesting to see that the ranking of effects changes with age. For the cross-sectional model at age 12 (a) 12-12 $f(\texttt{height})$ explains most of the variability in model (1); for all other age

combinations, (b) - (f) with the outcome log fat mass at ages 14 or 16, gender is the most important term, and dropping gender, halves the adjusted R^2. This has most likely to do with physiological changes through adolescence where the percentage body fat of girls increases compared to boys. Except for model (f), PA is the second most important predictor in terms of explaining variability.

4.1 Predicting fat mass with a redistributed PA

Using the 'best' model (1) we predict the percentage change in fat mass when taking away 15 minutes of sedentary activity (0 - 100 cpm) per day and redistributing this time evenly at intensities considered as moderate to vigorous (above 3600 cpm, scenario 1) and at intensities considered as vigorous (scenario 2). We compare these predictions with predictions from model (2), which assumes a linear effect of PA intensity. Model (2) is close to the current state-of-the-art and models similar to this have often been used in the past, see for example Riddoch et al. [28]. A comparison of confidence intervals from model (1) and model (2) shown in Figure (6) shows that model (1) always gives higher predictions of the percentage reduction in fat mass than model (2). Scenario 2 achieves a higher reduction in fat mass than scenario 1. A redistributed PA at age 12 according to scenario 2 achieves a predicted mean reduction of 14.3% of fat mass at age 16 (combination c) with a 95% credible interval between 12 - 17%.

5 Discussion

We introduce a novel modelling approach to model the effect of physical activity, as measured by accelerometer high-frequency count data, on fat mass or other health outcomes. We expect that our approach will be useful for other data of similar nature to the accelerometer time series. For example our approach could be applied to investigate the effect of daily pollution or temperature data on certain health outcomes. Our method can handle the challenges encountered with accelerometer and other high-frequency count data, including time series that are highly spiky and variable in nature, as well as of different lengths and with different patterns of missing values. The approach treats the accelerometer data as a function of activity intensity rather than time and hence avoids registration of accelerometer counts in time. It allows us to consider the entire distribution of accelerometer counts in the model without any of the collinearity problems introduced when summaries of specific activity levels are used simultaneously in the model. We show that leaving out the first histogram bin leads to the interpretation of the effect of activity at a certain intensity compared to sedentary activity. This parametrisation of the functional predictor in our model is similar in nature to the isotemporal substitution model [19], which estimates the effect of replacing one physical activity type with another physical activity type for the same amount of time.

We show how additive multi-dimensional functional predictors derived from accelerometer profiles can be used to check the assumptions intrinsically made by the use of commonly used single summary statistics of accelerometer profiles as predictors for a specific health outcome. In fact, in Section 4 we have shown that for the ALSPAC data the commonly made assumptions about the

effect of PA on log fat mass are not valid. We have also demonstrated how our approach can be used to answer specific questions about the effect of PA intensity on a particular outcome.

The model results are robust to changes in the number of bins. We selected the number of bins that minimised the prediction error (RMSE). When the number of bins was increased further model results became unstable. The histograms of PA are highly skewed with a spike at the first bin, but since we do not use the first bin in order to achieve model identifiability this is not a problem. We obtained very similar results from an analysis where we transform the scale of the counts by raising them to the power of 0.35. On this scale, an even spacing of bin widths results in a more uniform distribution, which is also more conducive to graphical display. Another possible transformation would be to use the logarithm. We decided to not transform the accelerometer counts in our main analysis in order to make the interpretation of results more straightforward. As model 1 contains model 2 with a linear effect function for the histogram as a special case, this also enables us to check whether the common modelling approach using the scalar summary statistic mean cpm is adequate.

Masse et al. [15] report on a sensitivity analysis comparing different processing protocols and conclude that the decision rules used to process accelerometer data have an impact on outcome variables and sample size. The processing protocol defines the length of a block of zeros for assuming non-weartime - resulting in setting those zeros to missing values. Here, changing block size from 10 zeros to 60 zeros to assess non-weartime makes little difference in the estimates. The analysis with 60 block zeros produces more or less the same graph as in Figure 5.

Our new functional modelling approach for accelerometer data opens the door to investigating many further questions of interest in the future. In particular, it would be interesting to develop a longitudinal model that allows the modelling of all visits simultaneously, although further work would be needed to investigate the optimal lag structure and possible collinearity when including more than one lag at a time. It would also be of interest to investigate whether the functional summaries can be used to classify individuals into groups with a similar activity pattern. Since there is evidence from many physiological studies that certain frequency patterns of PA at certain levels are beneficial, it would be of interest to confirm these findings with our modelling approach. Our method has potential to be useful for exploring such patterns. Applying our approach to other outcomes and/or high-frequency count data would also be of interest. The strength and shape of the relationship between PA and health outcome may differ by outcome. In particular we might expect to see different effects for the health outcomes blood pressure [13] or depression [35].

Instead of the histogram we could have used the kernel density estimate as a summary function. Repeating the analysis using kernel density estimates as a summary function gave similar answers. Both the histogram and the kernel density estimate summarise the distribution, with the kernel density estimate being smoother but also requiring user input, as the degree of smoothness has to be decided on by choosing a window width.

In ALSPAC PA levels were measured for a relatively short time, a maximum of seven days, and we assume that the measurements reflect the individuals' typical activity pattern. There are some exceptions as water based activities cannot be measured by the accelerometers used and some activities, such as cycling, are not well measured. The counts during water-based activities are

filtered out as non-weartime and set to missing. Ness et al. [23] showed that the cross-sectional association between PA and fat mass at age 12 was not affected if the analyses was repeated including only children who did not report swimming and cycling in the week of measurement. Given this evidence and the large sample size, we assume that these exceptions are negligible. The reliability and representativeness of accelerometer measurements is also supported by results from a repeated measures analysis [17], with individuals having repeats of wearing accelerometers. This study showed that the intra class coefficient (ICC) was moderately high for certain scalar summaries.

Although typical relative frequencies at high intensity levels are low, the time spent at the most intense levels of activity is very important. The estimates of $f(x_j)$ are negative and have overall the highest absolute values in that range, implying that apparently spending a certain number of minutes at the most intensive range of activity compared to spending time in sedentary activity is associated with the highest reduction of fat mass. This has implications regarding the effectiveness of public health interventions. Figure 6 shows that the potential benefit of redistributing 15 minutes of time per day from sedentary to moderate to vigorous activity is substantial (scenario 1) and a redistribution of time from sedentary to vigorous activity (scenario 2) has an even higher impact. An estimated mean reduction of 14.3% in fat mass at age 16 using model (1) with a credible interval of 12 - 17% when 15 minutes of time per day spent in sedentary activity at age 12 is redistributed to vigorous activity, is a modest yet clinically relevant effect. Redistributing say, 60 minutes, according to scenario 2 would be associated with a more substantial fat reduction. These predictions also show how our model can be used to investigate potential changes regarding PA. See also the approach of Mekary et al. [19] which allows similar investigations for questionnaire based physical activity data. In summary, our approach allows the investigation of the effect of PA intensity on a particular outcome in detail. It also allows the investigation of potential effects of a PA intervention and to optimise the impact of an intervention. This has implications regarding the effectiveness of public health interventions and PA intervention studies.

6 Supplementary Material

The supplementary material covers pre-processing of the accelerometer data and R example code.

Acknowledgements

References

[1] N. H. Augustin, M. Musio, K. von Wilpert, E. Kublin, S. N. Wood, and M. Schumacher. Modeling spatiotemporal forest health monitoring data. *Journal of the American Statistical Association*, 104(487):899–911, 2009.

[2] N. H. Augustin, C. Mattocks, A.R. Cooper, A.R. Ness, and J.J. Faraway. Modelling fat mass as a function of weekly physical activity profiles measured by actigraph accelerometers. *Physiological Measurement*, 33(11):1831–1839, 2012.

[3] A. Boyd, J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring, and G. D. Smith. Cohort profile: the children of the 90s - the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, 42(1):97–110, 2012.

[4] J. Ding, J. Symanzik, A. Sharif, J. Wang, S. Duntley, and W. D. Shannon. Powerful actigraphy data through functional representation. *Chance*, 24(3):30–36, 2011.

[5] P.H.C. Eilers and B.D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121, 1996.

[6] A. Fraser, C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey Smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S.M. Nelson, and D.A. Lawlor. Cohort profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42(1):97 – 110, 2013.

[7] J. Goldsmith, V. Zipunnikov, and J. Schrack. Generalized multilevel functional-on-scalar regression and principal component analysis. *Biometrics*, 2015.

[8] S. Greven and T. Kneib. On the behaviour of marginal and conditional aic in linear mixed models. *Biometrika*, 97(4):773–789, 2010.

[9] S. Grünewälder, F. Broekhuis, D. W. Macdonald, A. M. Wilson, J. W. McNutt, J. Shawe-Taylor, and S. Hailes. Movement activity based classification of animal behaviour with an application to data from cheetah (acinonyx jubatus). *PloS one*, 7(11):e49120, 2012.

[10] M. Grydeland, B. H. Hansen, M. Ried-Larsen, E. Kolle, and S. A. Anderssen. Comparison of three generations of actigraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports Science, Medicine and Rehabilitation*, 6(1):26, 2014.

[11] D.W. Haslam and W.P.T. James. Obesity. *Lancet*, 366(9492):1197–1209, 2005.

[12] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4): 963–974, 1982.

[13] S.D. Leary, A.R. Ness, G.D. Smith, C. Mattocks, K. Deere, S. N. Blair, and C. Riddoch. Physical activity and blood pressure in childhood: findings from a population-based study. *Hypertension*, 51(1):92–8, 2008.

[14] B.D. Marx and P.H.C. Eilers. Generalized linear regression on sampled signals and curves: A p-spline approach. *Technometrics*, 41:1–13, 1999.

[15] L. C. Masse, B. F. Fuemmeler, C. B. Anderson, C. E. Matthews, S. G. Trost, D. J. Catellier, and M. Treuth. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Medicine and Science in Sports and Exercise*, 37(11):S544, 2005.

[16] C. Mattocks, S. Leary, A. Ness, K. Deere, J. Saunders, K. Tilling, J. Kirkby, S. N. Blair, and C. Riddoch. Calibration of an accelerometer during free-living activities in children. *International Journal of Pediatric Obesity*, 2(4):218–226, 2007.

[17] C. Mattocks, S. Leary, A.R. Ness, K. Deere, J. Saunders, J. Kirkby, S. N. Blair, K. Tilling, and C. Riddoch. Intraindividual variation of objectively measured physical activity in children. *Medicine and Science in Sports and Exercise*, 39(4):622–629, 2007.

[18] C. Mattocks, A. Ness, S. Leary, K. Tilling, S.N. Blair, J. Shield, K. Deere, J. Saunders, J. Kirkby, G. Davey-Smith, J. Wells, N. Wareham, J. Reilly, and C. Riddoch. Use of accelerometers in a large field based study of children: Protocols, design issues, and effects on precision. *Journal of Physical Activity and Health*, 5:S94–S107, 2008.

[19] R. A. Mekary, W. C. Willett, F. B. Hu, and E. L. Ding. Isotemporal substitution paradigm for physical activity epidemiology and weight change. *American Journal of Epidemiology*, 170 (4):519–527, 2009.

[20] J. A. Mitchell, R. R. Pate, M. Dowda, C. Mattocks, C. Riddoch, A. R. Ness, and S. N. Blair. A prospective study of sedentary behavior in a large cohort of youth. *Medicine and Science in Sports and Exercise*, 44(6):1081, 2012.

[21] J.A. Mitchell, C. Mattocks, A.R. Ness, S.D. Leary, R.R. Pate, M. Dowda, S.N. Blair, and C. Riddoch. Sedentary behaviour and obesity in a large cohort of children. *Obesity (Silver Spring)*, 17(8):1596–1602, 2009.

[22] J.S. Morris, C. Arroyo, B.A. Coull, L. M. Ryana, R. Herricka, and S. L. Gortmakeret. Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles. *Journal of the American Statistical Association*, 101:1352–1364, 2006.

[23] A. Ness, S.D. Leary, C.G. Mattocks, S.N. Blair, J.J. Reilly, J. Wells, S. Ingle, K. Tilling, G. Davey Smith, and C. Riddoch. Objectively measured physical activity and fat mass in a large cohort of children. *PLoS Medicine*, 4 (3, e97):0476–0484, 2007.

[24] D.M. Pober, J. Staudenmayer, C. Raphael, and P.S. Freedson. Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and Science in Sports and Exercise*, 38:1626–1634, 2006.

[25] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

[26] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, New York, 2nd ed., 2005.

[27] J.J. Reilly, L.A. Kelly, D.M. Jackson, C. Montgomery, C. Slater, S. Grant, and J.Y. Paton. Validation of actigraph accelerometer estimates of total energy expenditure in young children. *International Journal of Pediatric Obesity*, 1(3):161–167, 2006.

[28] C.J. Riddoch, S. D. Leary, A. R. Ness, S. N. Blair, K. Deere, C. Mattocks, A. Griffiths, G. D. Smith, and K. Tilling. Prospective associations between objective measures of physical activity and fat mass in 12-14 year old children: the Avon Longitudinal Study of Parents and Children (ALSPAC). *British Medical Journal*, 339:b4544, 2009.

[29] M. Ried-Larsen, J. C. Brønd, S. Brage, B. H. Hansen, M. Grydeland, L. B. Andersen, and N. C. Møller. Mechanical and free living comparisons of four generations of the actigraph activity monitor. *Int J Behav Nutr Phys Act*, 9(113), 2012.

[30] F. Sera, M. Cortina-Borja, M. Geraci, L. Griffiths, C. Rich, and C. Dezateux. Modelling accelerometer data from 7-year old british children using function analysis of variance. *Journal of Epidemiology and Community Health*, 65:A26–A27, doi:10.1136/jech.2011.143586.59, 2011.

[31] H. Shou, V. Zipunnikov, C. M. Crainiceanu, and S. Greven. Structured functional principal component analysis. *Biometrics*, 71(1):247–257, 2014.

[32] J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107:1300–1307, 2009.

[33] G. Wahba. Bayesian confidence intervals for the cross validated smoothing spline. *J. R. Statist. Soc. B*, 45:133–150, 1983.

[34] J. Wang, H. Xian, A. Licis, E. Deych, J. Ding, J. McLeland, C. Toedebusch, T. Li, S. Duntley, and W. Shannon. Measuring the impact of apnea and obesity on circadian activity patterns using functional linear modeling of actigraphy data. *Journal of circadian rhythms*, 9(1):11, 2011.

[35] N.J. Wiles, A.M. Haase, D.A. Lawlor, A. Ness, and G. Lewis. Physical activity and depression in adolescents: cross-sectional findings from the ALSPAC cohort. *Soc Psychiatry Psychiatr Epidemiol*, 47(7):1023–1033, 2012.

[36] S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73 (1):3–36, 2011.

[37] S. N. Wood. A simple test for random effects in regression models. *Biometrika*, 100(4): 1005–1010, 2013.

[38] S.N. Wood. *Generalized Additive Models. An Introduction with R.* Chapman & Hall/CRC, Boca Raton, 2006.

[39] S.N. Wood, N. Pya, and B. Säfken. On doubly generalized additive models. *submitted*, 2014.

[40] L. Xiao, L. Huang, J. A. Schrack, L. Ferrucci, V. Zipunnikov, and C. M. Crainiceanu. Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics*, 67(2):352–367, 2015.

[41] M. Yang, H. Zheng, and H. Wang. A machine learning approach to assessing gait patterns for complex regional pain syndrome. *Medical Engineeeing & Physics*, 34(6):740–746, 2012.

[42] X. Yang, R. Telama, E. Leskinen, K. Mansikkaniemi, J. Viikari, and O. T. Raitakari. Testing a model of physical activity and obesity tracking from youth to adulthood: the cardiovascular risk in young Finns study. *International Journal of Obesity*, 31:521–527, 2006.

[43] S. Zhang, A. V. Rowlands, P. Murray, and T. Hurst. Physical activity classification using the genea wrist worn accelerometer. *Medicine and Science in Sports and Exercise*, 44:2228–2234, 2012.
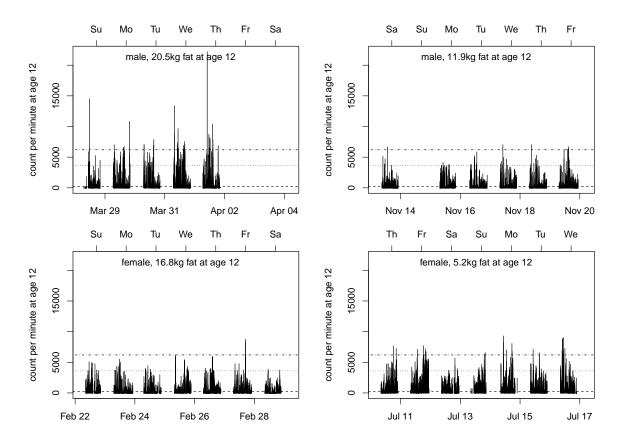
Funding

Figure 1: Sample of four PA profiles at age 12 years. The horizontal lines denote from bottom to top the cut points for light, moderate and vigorous PA at 200, 3600, 6200 counts per minute derived in the calibration study Mattocks et al. [16].
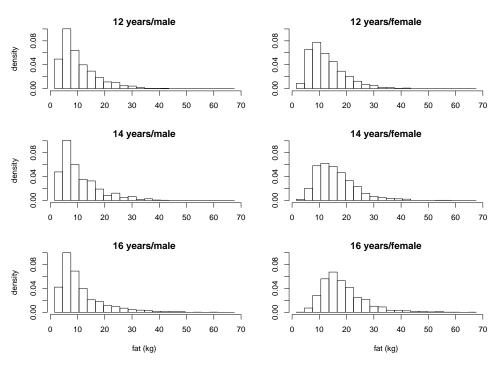
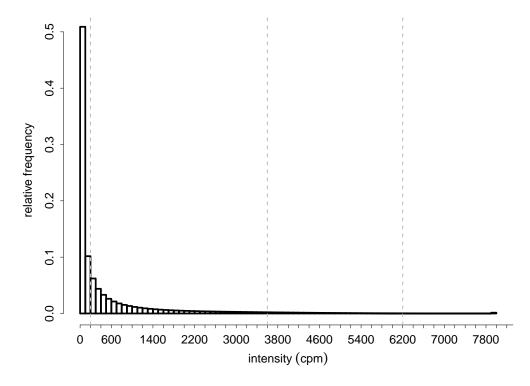Figure 2: Histogram of fat mass (kg) measured by scanner at ages 12, 14 and 16 for boys and girls.

Figure 3: The mean histogram (relative frequency) of PA profiles with 80 bins over all individuals and ages. The cut-points for light, moderate and vigorous PA currently used in ALSPAC [16] are at 200, 3600, 6200 counts per minute.
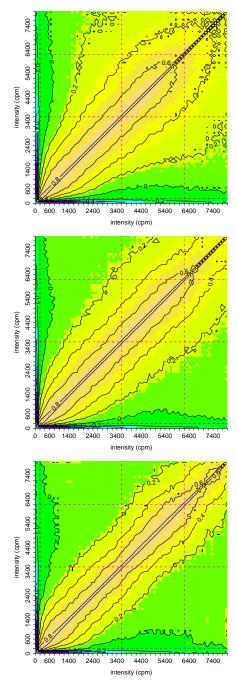
Figure 4: Estimated correlation between the density of the histogram with 80 bins for individuals at age 12 (top), 14 years (middle) and 16 (bottom). The dashed lines denote PA at 200, 3600, 6200 counts per minute. The labelled isolines denote the correlation. The colour scheme is from dark blue (correlation < -0.6), via green, yellow, orange (correlation > 0.6) to white (correlation = 1).
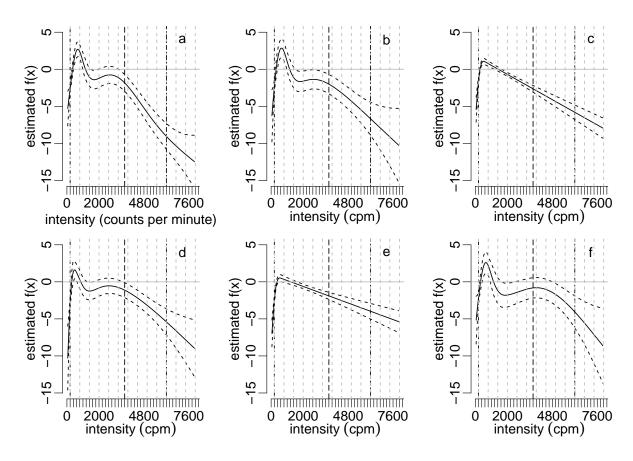
Figure 5: The estimated functions $f(p_j)$ for all six models. (a) 12-12 fat mass, height and weartime and PA at age 12; (b) 12-14 fat mass, height and weartime at age 14 and PA at age 12; (c) 12-16 fat mass, height and weartime at age 16 and PA at age 12; (d) 14-14 fat mass, height, weartime and PA at age 14; (e) 14-16 fat mass, height and weartime at age 16 and PA at age 14; (f) 16-16 fat mass, height, weartime and PA at age 16. The gray dashed vertical lines are at 200, 800, 1400, ..., 8000. The cut-points for light, moderate and vigorous PA from Mattocks et al. [16] estimated at 200, 3600, 6200 counts per minute are also shown as vertical lines.
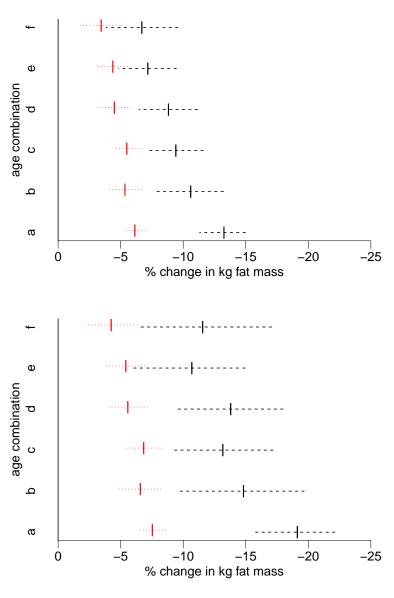
Figure 6: The 95% credible intervals for the percentage change in kg fat mass predicted for the best model (1) are shown with a black dashed line for each of the six different age combinations (a-f). The red dotted line shows the 95% credible intervals for the percentage change in kg fat mass predicted from model (2). Top: Scenario 1, predicted for a redistribution of 15 minutes activity evenly into activities between moderate to vigorous. Bottom: Scenario 2, predicted for a redistribution of 15 minutes sedentary (0 - 100 cpm) activity evenly into vigorous activities.

Table 1: Model selection results for models using a histogram with 80 bins. The models with different age combinations are fitted. We use the following short notations (a) 12-12, (b) 12-14,(c) 12-16,(d) 14-14, (e) 14-16 and (f)16-16, where the first number refers to the age of PA and weartime in the model and the second number refers to the age of the response fat mass and covariate height in the model. E.g. (b) 12-14 has PA and weartime at age 12 and response fat mass and covariate height at age 14. All models are compared to model 'base' ($\log(y_i) = x_i^T \boldsymbol{\beta} + f_h(height_i) + \epsilon_i$) which does not include a term for PA. The edf are the effective degrees of freedom of the model, R^2 is the adjusted R^2. For AIC and BIC the difference to the 'base' model is given. Parameters, adj. R^2, AIC and BIC are estimated from the training data with sample size N_T. The root mean squared prediction error (RMSPE) is in kg fat mass and estimated on the validation data.

age combination	model	N_T	edf	adj. R^2	Δ AIC	Δ BIC	RMSPE (kg)
(a) 12-12	base	2918	8	25.080	0.000	0.000	5.895
	+ hist (1)	2918	13	31.270	246.700	217.400	5.674
	+ cpm (2)	2918	9	28.060	117.100	110.900	5.841
	hist by gender (3)	2918	16	31.120	237.700	192.000	5.682
	2Dhist (4)	2918	28	31.810	255.000	138.100	5.706
	hist by WE (5)	2918	14	29.760	182.000	145.300	5.743
(b) 12-14	base	1736	6	23.330	0.000	0.000	6.625
	+ hist (1)	1736	12	28.050	105.200	77.000	6.442
	+ cpm (2)	1736	7	25.250	42.890	37.280	6.587
	hist by gender (3)	1736	14	27.790	96.180	53.180	6.416
	2Dhist (4)	1736	18	27.770	91.720	27.430	6.417
	hist by WE (5)	1736	12	26.300	63.610	35.940	6.468
(c) 12-16	base	1630	7	24.180	0.000	0.000	7.055
	+ hist (1)	1630	10	27.900	78.970	62.300	6.956
	+ cpm (2)	1630	8	25.980	37.940	32.110	7.062
	hist by gender (3)	1630	13	27.950	77.180	44.950	6.959
	2Dhist (4)	1630	20	27.190	53.370	-15.200	6.992
	hist by WE (5)	1630	11	26.230	40.720	19.140	7.047
(d) 14-14	base	2156	5	32.730	0.000	0.000	7.564
	+ hist (1)	2156	10	36.470	118.400	90.990	7.401
	+ cpm (2)	2156	6	34.960	71.550	65.880	7.548
	hist by gender (3)	2156	12	36.260	109.600	72.650	7.426
	2Dhist (4)	2156	18	36.320	105.700	34.320	7.405
	hist by WE (5)	2156	11	35.760	93.060	58.240	7.470
(e) 14-16	base	1656	5	32.090	0.000	0.000	7.852
	+ hist (1)	1656	9	35.340	77.340	57.310	7.824
	+ cpm (2)	1656	6	33.720	39.020	33.620	7.873
	hist by gender (3)	1656	11	35.200	71.580	39.930	7.839
	2Dhist (4)	1656	14	33.560	27.450	-20.210	7.836
	hist by WE (5)	1656	9	34.630	58.680	35.300	7.862
(f) 16-16	base	963	5	33.940	0.000	0.000	8.048
	+ hist (1)	963	8	34.810	9.750	-5.228	8.011
	+ cpm (2)	963	6	34.420	6.101	1.232	8.018
	hist by gender (3)	963	11	35.590	18.430	-10.880	7.899
	2Dhist (4)	963	13	35.150	10.110	-27.770	7.954
	hist by WE (5)	963	9	34.860	9.403	-10.530	7.843

Table 2: The effect of dropping each term in turn from model (1) for the six age combinations (a) - (f) is shown. edf are the effective degrees of freedom of the model, R.sq is the adjusted R^2. Parameters, adj. R^2, AIC and BIC are estimated from the training data with sample size N_T. For AIC and BIC the difference to the 'base' model is given. The root mean squared prediction error (RMSPE) is estimated in the validation data.

age comb.	model	N_T	edf	adj. R^2	Δ AIC	Δ BIC	RMSPE
(a) 12-12	model (1)	2918	13	31.2700	0.0000	0.0000	5.6740
	-gender	2918	12	29.8600	-58.4800	-53.8400	5.7250
	-m.obese	2918	12	29.0600	-91.7300	-86.2600	5.7470
	-weartime	2918	12	31.1800	-3.1010	2.7690	5.6720
	$-\sum_j f(p_j)z_i(p_j)$	2918	8	25.0800	-246.7000	-217.4000	5.8950
	$-f_h(\texttt{height}_i)$	2918	9	15.1000	-612.5000	-587.5000	6.1560
(b) 12-14	model (1)	1736	12	28.0500	0.0000	0.0000	6.4420
	-gender	1736	12	16.8100	-251.9000	-251.7000	6.9110
	-m.obese	1736	11	25.7400	-53.8100	-48.0100	6.5230
	weartime	1736	11	27.9500	-1.5480	3.7820	6.4540
	$-\sum_j f(p_j)z_i(p_j)$	1736	6	23.3300	-105.2000	-77.0000	6.6250
	$-f_h(\texttt{height}_i)$	1736	9	24.6200	-78.3600	-64.4100	6.6110
(c) 12-16	model (1)	2156	10	36.4700	0.0000	0.0000	7.4010
	-gender	2156	13	16.4700	-592.8000	-608.5000	8.1270
	-m.obese	2156	9	34.5200	-64.0000	-57.6100	7.5850
	weartime	2156	9	36.3600	-2.7200	3.2220	7.4010
	$-\sum_j f(p_j)z_i(p_j)$	2156	5	32.7300	-118.4000	-90.9900	7.5640
	$-f_h(\texttt{height}_i)$	2156	9	33.5600	-95.3400	-88.4800	7.5340
(d) 14-14	model (1)	1630	10	27.9000	0.0000	0.0000	6.9560
	-gender	1630	12	15.4500	-262.0000	-273.4000	7.2830
	-m.obese	1630	9	25.3700	-55.2000	-48.7500	7.0950
	-weartime	1630	9	27.6600	-4.5150	0.9018	6.9220
	$-\sum_j f(p_j)z_i(p_j)$	1630	7	24.1800	-78.9700	-62.3000	7.0550
	$-f_h(\texttt{height}_i)$	1630	7	25.2300	-56.3800	-40.1300	7.1070
(e) 14-16	model (1)	1656	9	35.3400	0.0000	0.0000	7.8240
	-gender	1656	13	15.1900	-453.1000	-474.9000	8.6350
	-m.obese	1656	8	33.1000	-55.3200	-49.7000	8.0030
	-weartime	1656	7	35.1700	-2.9810	3.7420	7.8060
	$-\sum_j f(p_j)z_i(p_j)$	1656	5	32.0900	-77.3400	-57.3100	7.8520
	$-f_h(\texttt{height}_i)$	1656	8	32.8200	-62.2900	-57.6700	7.9590
(f) 16-16	model (1)	963	8	34.8100	0.0000	0.0000	8.0110
	-gender	963	9	13.4800	-273.6000	-278.9000	8.2790
	-m.obese	963	9	33.0300	-26.9500	-32.1500	7.9400
	-weartime	963	8	34.9700	2.7560	4.8210	7.9860
	$-\sum_j f(p_j)z_i(p_j)$	963	5	33.9400	-9.7500	5.2280	8.0480
	$-f_h(\texttt{height}_i)$	963	9	33.6500	-17.8600	-22.4800	7.9080