Estimation of Exercise Energy Expenditure Using a Wrist-Worn Accelerometer: a Linear Mixed Model Approach with Fixed-Effect Variable Selection

Eija Haapalainen[1], Perttu Laurinen[1], Juha Röning[1], Hannu Kinnunen[2]

[1] Intelligent Systems Group, P.O. BOX 4500, FI-90014 University of Oulu, Finland
{Eija.Haapalainen, Perttu.Laurinen, Pekka.Siirtola, Juha.Roning}@ee.oulu.fi

[2] Polar Electro Oy, Finland

Abstract

This article presents an approach to estimating exercise energy expenditure based on acceleration measurements from a wrist-worn biaxial sensor. The method uses the linear mixed model that makes it possible to model both between-subject and within-subject variation in energy expenditure. More precisely, a random-intercepts model is used. The variance and mean of the acceleration signals at 15-second intervals as well as subject demographics (height, weight, body mass index, age and VO_2max) are used. Energy expenditure is modelled in four different activities: walking, running, Nordic walking and bicycling. This study introduces an effective backward model selection procedure for selecting the fixed-effect variables in the model. The procedure uses leave-one-out cross-validation to be able to effectively exploit the available data set and to ensure the robustness of the model. Estimation accuracy in test sets is used as a criterion of model performance. The model selection procedure proposed notably improves estimation accuracy. In walking, running, Nordic walking and bicycling, average estimation errors of 3.9, 3.6, 1.9 and 13.5 percent are reached. The respective Pearson correlations for these activities are 0.91, 0.98, 0.97, and 0.81. These results are also compared to the performance of the general linear model. It is discovered that the linear mixed model outperforms the model that does not take the individual levels of energy expenditure of the subjects into account.

1 Introduction

There are more than 1 billion overweight adults in the world and at least 300 million of them obese. These conditions pose a major risk for chronic diseases, including type 2 diabetes, cardiovascular disease, hypertension and stroke, and certain forms of cancer [15]. Reduced physical activity is one of the main reasons for this problem. Therefore, development of objective and reliable methods for accessing physical activity is highly important.

Energy expenditure caused by physical activity is commonly accepted as the standard reference of physical activity in humans [6]. Energy expenditure can be reliably measured from a person's oxygen consumption. However, measurement of oxygen consumption requires the use of a breath gas analyzer (indirect calorimetry) and is therefore impractical and not feasible under free-living conditions.

Research on modelling physical activity in different activities based on acceleration data has expanded over the past two decades [2], [12]. Placement and orientation of body-mounted accelerometers have been studied by several authors. Sensors attached to the low back, near the centre of gravity, have been commonly used [1], [10], [11], [14]. However, Mathie *et al* [7] point out that waist-mounted accelerometers significantly underestimate activity concentrated in the upper body. Therefore, in some studies additional acceleration sensors have been attached to other body parts such as chest, wrist, hip, thigh or ankle (*e.g.* [10]).

In most of the previous studies, regression methods have been applied to accelerometer counts and oxygen consumption simultaneously measured to determine the relationship between the two measures and to define an equation to predict energy expenditure from acceleration [12]. The accelerometer counts are usually obtained by integrating the accelerometer signal [3]. Rothney *et al* [10] presented an artificial neural network approach to estimating energy expenditure based on features extracted from raw accelerometer data and subject demographics.

A previous study by the authors [5] introduced the idea of modelling energy expenditure based on acceleration measurements using the linear mixed model. Instead of the accelerometer counts, the variance of the raw accelerometer signal at 15-second intervals was used. Oxygen consumption was modelled based on measurements from a wrist-worn biaxial accelerometer. This location was selected since it is an easy and comfortable location for the subject and does not disturb performance of the activity. In walking, running and Nordic walking the model underestimated total energy expenditure by 13, 2 and 9 percent, respectively, and in bicycling total energy expenditure was overestimated by 7 percent. In that study, however, the performance of the model was only tested on the measurements of one subject not included in the training set.

The aim of this study is two-fold. First, the usefulness of the mixed model-based approach is further validated and new covariates are introduced to the model to improve estimation accuracy. Secondly, an efficient model selection procedure is developed to more efficiently exploit the available data set and to ensure the robustness of the model.

Verbeke and Molenberghs [13] mention approximate Wald tests, t-tests and F-tests, likelihood ratio tests and the frequently used information criteria AIC, SBC, HQIC and CAIC as means of selecting the fixed effect variables in a linear mixed model. The Akaike Information Criterion (AIC) have been used by Ngo and Brand [8], and the Generalized Information Criterion (GIC) by Pu and Niu [9]. Fernandez [4] used both the Corrected Akaike Information Criterion (AICC) and the Minimal Description Length (MDL). However, the weakness of all these methods is that they go through all possible models and are therefore computationally expensive. Verbeke and Molenberghs [13] also warn about using information criteria to discriminate between several statistical models and advise not to interpret them as formal statistical tests of significance.

In this study a greedy backward model selection procedure is presented for fixed-effect variable selection in linear mixed models. The method uses leave-one-out cross-validation to select the best model. A t-test is used to measure the contribution of each term in the presence of other variables in the model.

2 Methods

2.1 The Linear Mixed Model

The *linear mixed model* is very suitable for modelling time series data such as oxygen consumption containing repeated measures from several subjects. The repeated measurements are correlated with each other, which has to be taken into account in the modelling.

Classical statistics assume that observations are independent and identically distributed. The linear mixed model, however, assumes two sources of variation: subject-specific and population-specific. The vector of repeated measurements of each subject is assumed to follow a linear regression model where some of the regression parameters are common for all the subjects, whereas other parameters differ between subjects [13]. Therefore, the model offers the possibility to provide each of the subjects with an individual ground level of energy expenditure.

The linear mixed model is of the form

$$\begin{aligned} &Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \\ &b_i \overset{i.i.d}{\sim} N(0, D), \\ &\varepsilon_i \overset{i.i.d}{\sim} N(0, \Sigma_i), \quad 1 \leq i \leq N, \end{aligned}$$

where N is the number of subjects, Y_i is the n_i-dimensional vector of observations for subject i, X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional fixed matrices of known covariates, β is a p-dimensional unknown fixed-effects parameter vector, b_i is the q-dimensional vector of random effects, and ε is an n_i-dimensional vector containing the within-subject error components. D is a general symmetric $(q \times q)$ covariance matrix and Σ_i is a $(n_i \times n_i)$ covariance matrix that has the property that the set of unknown parameters in Σ_i do not depend upon i. [13]

2.2 Model Structure

An overall intercept term was fitted to model the average level of oxygen consumption, and subject-specific intercepts to model the individual ground level of each of the subjects. No other subject-specific effects were specified. This is the so-called *random-intercepts model* [13], where the regression coefficients are the same for each subject. When the model is applied to measurements outside the data set used to train it, no information on the subject-specific intercepts is available. Therefore, only the overall intercept term and the fixed regression structure are used in energy expenditure estimation.

The covariance structure Σ_i for the error components ε_i was specified to be first-order autoregressive (AR(1)). The maximum likelihood estimation method was used to estimate the covariance parameters.

2.3 Model Selection

A specific backward model selection procedure was developed for model selection in this application. It is based on selection of the fixed-effect variables used in the model by iterating training of the model and testing its performance until the best model structure is found. Leave-one-out cross-validation is used to test the model and choose the explanatory variables to be used.

The search for the most parsimonious model structure is started by fitting a model with all the explanatory variables available. After that one variable at a time is excluded from the model. The term to be excluded is chosen based on the

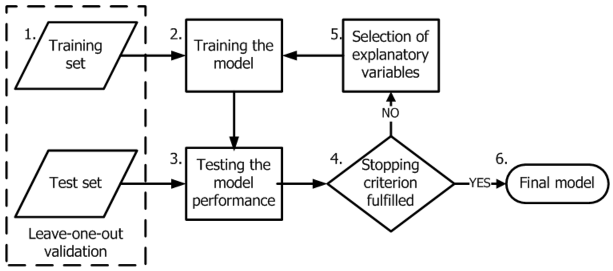

Figure 1. Backward model selection procedure using leave-one-out cross-validation

	Men (n=8)	Women (n=2)
Age (years)	29.8 ± 4.7 (22-37)	24.5 ± 4.9 (21-28)
Height (cm)	181 ± 5.6 (169-188)	164 ± 1.4 (163-165)
Body mass (kg)	82.6 ± 13.1 (62-104)	56.5 ± 7.8 (51-62)
BMI (kg/m^2)	25.1 ± 3.4 (21.6-31.1)	21.0 ± 2.5 (19.2-22.8)
VO_2max (ml/kg/min)	51.0 ± 13.7 (36-76)	49.5 ± 2.1 (48-51)

Table 1. Physical characteristics of the subjects (mean, standard deviation, range)

significance probability of the fixed-effects in the model. The exclusion of the variables is continued until a given stopping criterion is fulfilled. The functionality of the procedure used is presented in Figure 1.

First, the data set is divided into leave-one-out cross-validation data sets. Thus, there are as many training set-test set pairs as there are subjects in the data. These sets are marked with the number 1 and outlined with a dashed rectangle in Figure 1. The model selection is started by fitting the linear mixed model to each of the training sets (step 2). In step 3 the performance of the model is estimated using the test sets. The total prediction error of the iteration is defined as the average of the errors in the separate test sets. On each iteration a check of whether or not a predetermined stopping criterion is fulfilled is performed (step 4).

If the stopping criterion is not met, one fixed-effect variable is excluded from the model (step 5). The term to be dropped is selected on the grounds of the significance probability of the variables. The p-values for the t-tests measure the contribution of each explanatory variable in the presence of all other variables in the model. The fixed effect whose average p-value in the separate models fitted to each of the training sets is largest is excluded from the model, presuming that this value is greater than a prespecified threshold value. The use of average significance probability is possible since all the models have the same structure.

Iteration of steps 2-5 is continued until the stopping criterion is met. The stopping criterion consists of two alternative conditions. The iteration is stopped if either the maximum of the average p-values of the fixed effects in the models is less than the threshold value, or if there are only one fixed-effect variable left in the model. Finally, the model structure that has the smallest average estimation error in the leave-one-out test sets is selected as the final model. Execution of the procedure ends at step 6.

In this study, the model selection procedure presented is applied to selection of fixed-effect variables in a random-intercepts model. However, the method can also be applied to selection of fixed effects variables in any general linear mixed model presuming that all the models compared involve the same random effects. In addition, explanatory variables in a general linear model can be selected using the method as is done in the next Section where the linear mixed model performance is compared to the general linear model.

3 The study

3.1 Experimental Protocol

The data were collected from ten healthy subjects (eight men and two women). The physical characteristics of the participants are shown in Table 1. Body mass index (BMI) was calculated by dividing the body mass (kg) by the height squared (m^2). The maximal oxygen consumption of the subjects (VO_2max) was measured using indirect calorimetry in treadmill running until voluntary exhaustion.

The participants performed four different activities: *walking*, *running*, *Nordic walking* and *bicycling*. Each of the subjects performed one ten-minute test in each of the activities. Before the test period the subjects performed a warm-up for five minutes in order to raise the level of oxygen consumption before the actual testing period. However, in spite of the warm-up period, the oxygen consumption level ascended steeply during the first minute as the participants began the exercising. Therefore, the first minute of measurements was not included in modelling.

3.1.1 Measurement of Energy Expenditure

Oxygen consumption VO_2 (ml/min) of the subject was measured during the exercise using indirect calorimetry. Breath-by-breath data were collected using the Cosmed K4b^2 breath gas exchange measurement system. The K4B^2 is a light portable gas analyzer that uses a face mask. The concentration of expiration gases was measured at intervals of 15 seconds. The effect of the weight of the subject was taken into account by dividing the values of oxygen consumption by the weight of the subject hence converting the unit of measurement to ml/kg/min.

3.1.2 Measurement of Acceleration

The physical activity of the subject was measured using a biaxial accelerometer worn on the left wrist. The measuring device was composed of two 1-dimensional capacitive accelerometers (VTI Technologies) positioned perpendicular to each other. The sampling frequency was 100 Hz. The acceleration signals were preprocessed by replacing erroneous values falling outside the measurement range of the accelerometer (± 1.5 g) by the mean of the signal values.

3.1.3 Data Sets

The bicycling data contain measurements from nine subjects (seven men, two women), the walking and Nordic walking data are comprised of seven tests (six men, one woman) and the running data consist of measurements from six of the participants (five men, one woman). Although all ten participants performed the four activities, the measurement data could not be received from all the activities due to technical problems with the accelerometer data logger.

3.2 Feature extraction

The variance and mean of the absolute value of the original acceleration signals were calculated at intervals of 15 seconds. These features present the information contained in the original signals measured at a frequency of 100 Hz in a more compact form and also focus the acceleration measurements to the measurements of oxygen consumption made every 15 seconds.

3.3 Implementation of the Model

Oxygen consumption was set as the response variable in the model. The covariates include the variances and the means of the two acceleration signals, the interaction terms between the two variance variables and between the two mean variables, respectively, as well as information on the physical characteristics of the subjects. The subject demographics include height, weight, body mass index, age and the maximal oxygen consumption of the subject. In addition to the acceleration measurements made simultaneously to the oxygen consumption measurements, five lagged values of acceleration were also used. This means the influence of acceleration from 90 seconds preceding the oxygen consumption measurement were taken into account. The model gives an estimate of the oxygen consumption of the subject at 15-second intervals. The model was fitted using the SAS MIXED procedure (SAS Release 9.1.3.).

For each of the four activities, the optimal linear mixed model structure was found using the model selection procedure described in Subsection 2.3. All the explanatory variables presented above were used in the first iteration of the procedure and after that the less significant fixed-effect variables were dropped from the model one by one. The threshold value of 0.05 was used for the significance of the effects. The prediction error was estimated by comparing the total energy expenditure calculated from the oxygen consumption measurement and the respective estimate.

3.4 Results

The model was tested by the measurements of each of the subjects in the data set in turn. The measurements of one of the subjects were excluded from the data set and the model was trained using the measurements of the other subjects. The energy expenditure of the one subject constituting the test set was then estimated. This was repeated for all the subjects.

The modelling results in the four activities (walking, running, Nordic walking and bicycling) are shown in Figures 2 a) - d). The vertical axis represents the average energy expenditure during the nine-minute exercise calculated from the oxygen consumption measured using indirect calorimetry. The value on the horizontal axis is the energy expenditure calculated from the estimate given by the linear mixed model. Energy expenditure is here presented in units of metabolic equivalents, MET. It is the ratio of the working metabolic rate of a person relative to the resting metabolic rate. One MET equals an energy expenditure of 3.5 ml/kg/min. Thus, estimated oxygen consumption (VO_2/kg) is converted to METs by dividing by 3.5.

It can be seen from Figures 2 a) - d) that the model estimates energy expenditure very accurately. For walking, running and Nordic walking, the correlation between estimated and measured energy expenditure is almost perfectly linear. For bicycling, the group of markers is more scattered but the results are nevertheless very good. The Pearson correlation coefficients between measured and estimated energy expenditure in walking, running and Nordic walking are 0.91, 0.98 and 0.97, respectively, and 0.81 in bicycling, as seen in Table 2.

The average estimation errors of energy expenditure in the different activities are shown in Table 2. In walking, running, Nordic walking and bicycling the average estimation errors are 3.9, 3.6, 1.9 and 13.5 percent, respectively. In walking, running and Nordic walking the arms of the person perform an intense cyclic movement. In cycling, however, the arms are mostly immobile. Therefore, the fact that it is more difficult to estimate energy expenditure in bicycling based on measurements from a wrist-worn accelerometer feels natural.

There were altogether 31 covariates from which the backward model selection procedure presented in Section 2 selected the final set of variables to be used in the model. For each of the activities, a distinct set of variables was se-

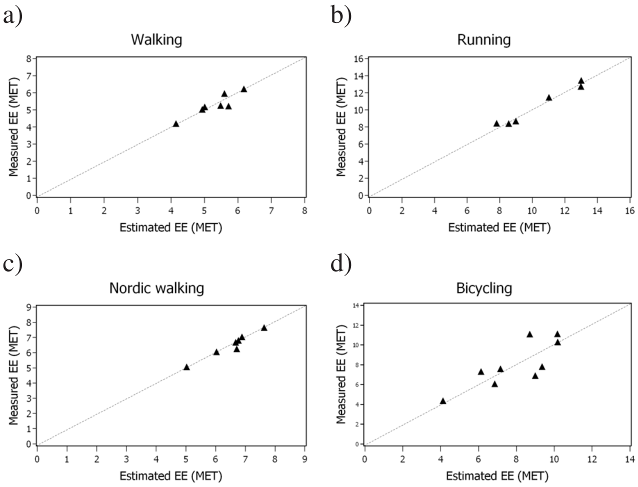

Figure 2. Estimation results of energy expenditure (EE) in four different activities. a) Walking b) Running c) Nordic walking d)Bicycling

		Walking	Running	Nordic walking	Bicycling
	No. of subjects	7	6	7	9
	No. of variables	3	30	17	9
GLM	**r**	0.87	0.96	0.91	0.75
	Estimation error	11.1 %	7.6 %	5.7 %	13.9 %
	No. of variables	4	12	13	3
LMM	**r**	0.91	0.98	0.97	0.81
	Estimation error	3.9 %	3.6 %	1.9 %	13.5 %
LMM	**r**	0.42	0.96	0.11	0.55
all variables	**Estimation error**	206.1 %	7.6 %	27.7 %	32.6 %

Table 2. Number of subjects and explanatory variables, correlation, and estimation error: general linear model, linear mixed model and linear mixed model without model selection

lected. The number of explanatory variables in the models for different activities varies considerably. For walking, running, Nordic walking and bicycling the models have 4, 12, 13 and 3 fixed effect variables, respectively. The model selection procedure significantly improves the accuracy of estimation. If all the explanatory variables available were used in the linear mixed model with no selection at all, the estimation errors for the same activities would be 206.1, 7.6, 27.7 and 32.6 percent. Hence, the model selection procedure presented in this study indeed notably improves the model's accuracy.

Also a general linear model with no subject-specific intercepts to model the individual ground levels of oxygen consumption was fitted to the data using the same fixed-effect variable selection method. The average estimation errors in the four activities were 11.1, 7.6, 5.7 and 13.9 percent, hence being inferior to the results obtained using the linear mixed model.

3.5 Discussion

The results of the linear mixed model-based approach to exercise energy estimation presented in the preceding section are excellent. Compared with the estimation errors in the same activities reported in the previous article by the authors [5], the results have improved substantially. In addition, the results have now been further validated by testing the model's performance using leave-one-out cross-validation and comparing the estimation accuracy to that of the general linear model.

In this study, a special model selection procedure for selecting fixed effect variables in a linear mixed model was presented. The method is a backward procedure in which leave-one-out cross-validation is used to effectively exploit the available data set and to minimize the effect of between-subject variation in energy expenditure. Significance probability of the fixed effects is used to measure the contribution of the explanatory variables in the model. The presented method makes model selection in linear mixed models automated and efficient. It can also be applied to selection of explanatory variables in general linear models.

Different information criteria are often used in model selection to discriminate between several models. Interpretation of the information criteria is, however, not straightforward and there are problems in using them in model selection. In this study the estimation accuracy of the model in a test set was used as the criterion of goodness of the model. Leave-one-out validation was used in testing the performance of the candidate models.

Several methods have been proposed for model selection in linear mixed models. The methods of Ngo and Brand [8], Pu and Niu [9] and Fernandez [4] are based on the use of different information criteria in comparing candidate models. However, the weakness of all these methods is that they rely on an exhaustive search of all the possible models. This makes the methods computationally very expensive if the number of available effects is large. The procedure presented in this paper is based on backward selection of the fixed effect variables. Hence, not all the possible models are tested. Therefore, the method has an evident computational advantage. There were altogether 31 available explanatory variables in this study. An exhaustive search of all possible effect combinations would have required evaluation of $2^{31} - 1 > 2 \cdot 10^9$ candidate models. Using the backward selection method, only 31 models were evaluated. Nevertheless, excellent modelling results were obtained.

A backward method for model selection was selected because of its easy implementation. Even though the procedure only explores a small proportion of all the possible models, the results of this study show that, combined with cross-validation, the backward method is an effective choice for selecting linear mixed models. The cross-validation approach could also be applied with other more complex model selection algorithms.

Compared with the methods that use different information criteria, the weakness of the method presented in this study could be claimed to be that it does not penalize for a large number of effects in the model. However, as the number of available effects in this application is relatively small and the procedure seems to select models where the number of fixed effects is small or at most moderate, this is not a problem.

In this study energy expenditure was modelled based on acceleration measurements from a biaxial wrist-worn sensor. The same approach could of course be applied to acceleration measured from other body parts. Estimation accuracy would probably be even improved if multiple sensors were used. However, for the convenience of the user and ease of use of the energy expenditure monitor, only one wrist-worn sensor was chosen to be used in this study.

4 Conclusions and Future Work

In this study, the use of a linear mixed model in estimating exercise energy expenditure based on acceleration measurements from a biaxial wrist-worn sensor was discussed. A unique model selection procedure for selecting the model structure was introduced. The procedure is based on backward selection of the fixed effects in the model by iterating training of the model and testing its performance until the optimal model structure is found. Leave-one-out cross-validation is used to test the model and choose the covariates to be used. The selection of variables is done on the grounds of significance probability of the effects in the model. The procedure enables efficient exploitation of the available data set in selecting the model structure and ensures robustness of the model by minimizing the effect of between-subject variation in energy expenditure.

Excellent results were obtained in modelling energy expenditure in four different activities. However, in this study the model selection was optimized to the data set available. For the performance of the model selection procedure to be further validated it should be applied to larger data sets with separate validation sets. The presented approach could also be applied to measurements from other activities. In addition, other more complex model selection algorithms than the backward method used in this study could be used.

5 Acknowledgements

The authors would like to thank Heidi Jurvelin who collected the data. In addition, we would like to express our gratitude for the financial support provided by the Graduate School in Electronics, Telecommunications and Automation, the Foundation of Technology, the Tauno Tönning Foundation and the Nokia Foundation.

References

[1] C. Bouten, A. Sauren, M. Verduin, and J. Janssen. Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Med Biol Eng Comput*, 35(1):50–6, 1997.

[2] C. Bouten, K. Westerterp, M. Verduin, and J. Janssen. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Med Sci Sports Exerc*, 26:1516–1523, 1994.

[3] K. Chen and D. Basset Jr. The technology of accelerometry-based activity monitors: Current and future. *Med Sci Sports Exerc*, 37(11 Suppl):S490–S500, 2005.

[4] G. Fernandez. Model selection in proc mixed - a user-friendly sas macro application. In *SAS Global Forum*, 2007.

[5] E. Haapalainen, P. Laurinen, P. Siirtola, J. Röning, H. Kinnunen, and H. Jurvelin. Exercise energy expenditure estimation based on acceleration data using the linear mixed model. In *IEEE IRI*, pages 131–136, 2008.

[6] R. LaPorte, H. Montoye, and C. Caspersen. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep*, 100(2):131–46, 1985.

[7] M. Mathie, A. Coster, N. Lovell, and B. Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement*, 25(2):R1–R20, 2004.

[8] L. Ngo and R. Brand. Model selection in linear mixed effects models using sas proc mixed. In *SUGI 22*, pages 1–6, 1997.

[9] W. Pu and X.-F. Niu. Selecting mixed-effects models based on a generalized information criterion. *Journal of Multivariate Analysis*, 97(3):733–758, 2006.

[10] M. Rothney, M. Neumann, A. Beziat, and K. Chen. An artificial neural network model of energy expenditure using non-integrated acceleration signals. *J Appl Physiol*, 103:1419–1427, 2007.

[11] P. Terrier, K. Aminian, and Y. Schutz. Can accelerometry accurately predict the energy cost of uphill/downhill walking? *Ergonomics*, 44(1):48–62, 2001.

[12] R. Troiano. Translating accelerometer counts into energy expenditure: advancing the quest. *J Appl Physiol*, 100(4):1107–1108, 2006.

[13] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, New York, 2000.

[14] G. Welk, J. Schaben, and J. Morrow Jr. Reliability of accelerometry-based activity monitors: a generalizability study. *Med Sci Sports Exerc*, 36(9):1637–45, 2004.

[15] WHO. Obesity and overweight, 2003. www.who.int/dietphysicalactivity/publications/facts/obesity (25 June, 2008).