

Validation of accelerometer for measuring physical activity in free-living individuals

Authors' Contribution:

A Study Design
B Data Collection
C Statistical Analysis
D Data Interpretation
E Manuscript Preparation
F Literature Search
G Funds Collection

George O. Agogo^{1, 2 ACDEF}, Hilko van der Voet^{1 DE}, Paul J.M. Hulshof^{3 AB}, Pieter van 't Veer^{3 DE}, Laura Trijsburg^{3, 4 ABCDEF}, Fred A. van Eeuwijk^{1 DE}, Hendrick C. Boshuizen^{1, 3, 5 ACDEF}

¹ Dep. of Biometrics, Wageningen University & Research, Wageningen, The Netherlands

² Dep. of Internal Medicine, Yale University, New Haven, USA

³ Div. of Human Nutrition, Wageningen University & Research, Wageningen, The Netherlands

⁴ Nutritional Methodology and Biostatistics Group, International Agency for Research on Cancer, Lyon, France

⁵ Dep. of Statistics, Mathematical Modelling and Data Logistics, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

abstract

Background: The aim of this research was to validate a triaxial GT3X accelerometer against doubly labelled water for measuring total energy expenditure (TEE) in a study of free-living Dutch adults and to compare the two prediction equations used to calculate accelerometer-derived activity related energy expenditure.

Material/Methods: We used a measurement error model to estimate bias in the mean TEE, a correlation coefficient between measured and true TEE (a validity coefficient, which quantifies loss of statistical power to detect association) and the attenuation factor (which quantifies bias in the association), with and without conditioning on age, sex and BMI. We proposed a calibration method for the accelerometer-based TEE.

Results: The accelerometer underestimated TEE by about 500kcal/day. The validity coefficient estimate conditional on age, sex and BMI was 0.8; the same value was observed for the attenuation factor estimate. With the devised calibration method, the bias in accelerometer-derived mean TEE reduced to 6 kcal/day, validity coefficient estimate increased to 0.95 and attenuation factor to 0.94.

Conclusions: The GT3X accelerometer would underestimate mean TEE, lead to minimal loss in statistical power to detect significant association, and would result in biased estimate of the association between TEE and a health outcome.

Key words: attenuation, measurement error, physical activity, activity energy expenditure, validity coefficient.

article details

Article statistics: Word count: 4,637; Tables: 5; Figures: 2; References: 33

Received: August 2017; **Accepted:** November 2017; **Published:** March 2018

Full-text PDF: <http://www.balticsportscience.com>

Copyright © Gdansk University of Physical Education and Sport, Poland

Indexation: Celdes, Clarivate Analytics Emerging Sources Citation Index (ESCI), CNKI Scholar (China National Knowledge Infrastructure), CNPIEC, De Gruyter - IBR (International Bibliography of Reviews of Scholarly Literature in the Humanities and Social Sciences), De Gruyter - IBZ (International Bibliography of Periodical Literature in the Humanities and Social Sciences), DOAJ, EBSCO - Central & Eastern European Academic Source, EBSCO - SPORTDiscus, EBSCO Discovery Service, Google Scholar, Index Copernicus, J-Gate, Naviga (Softweco, Primo Central (ExLibris), ProQuest - Family Health, ProQuest - Health & Medical Complete, ProQuest - Illustrata: Health Sciences, ProQuest - Nursing & Allied Health Source, Summon (Serials Solutions/ProQuest, TDOne (TDNet), Ulrich's Periodicals Directory/ulrichsweb, WorldCat (OCLC)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interests: Authors have declared that no competing interest exists.

Corresponding author: Dr. George O. Agogo; Wageningen University & Research, Wageningen, The Netherlands; e-mail: george.agogo@yale.edu; phone number: +1 203-737-2908

Open Access License: This is an open access article distributed under the terms of the Creative Commons Attribution-Non-commercial 4.0 International (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non-commercial and is otherwise in compliance with the license.

INTRODUCTION

Regular physical activity (PA) is associated with good health and greater degree of independence [1, 2]. Low levels of PA, however, are shown to be associated with diseases such as cardiovascular disease, type 2 diabetes, depression, osteoporosis, obesity and colon/breast cancer [1]. Sedentary lifestyle is a major concern to public health and is a health risk factor [3]. Physical activity involves body movement produced by skeletal muscles resulting in energy expenditure above resting levels [4, 5] and is a complex construct that involves sports and non-sports activities. The health benefits associated with regular PA are assessed by considering an individual's long-term average physical activity level over a long period of time (hereafter, usual activity) [6]. To correctly measure PA, it is important to monitor PA patterns (duration, frequency and intensity) and activity related energy expenditure. Ideally, usual activity would be measured without error in free-living individuals. In practice, however, it is difficult to measure PA without error under free-living conditions. Thus, usual activity measurements are subject to error, because some aspects of physical activity may not be captured.

Physical activity contributes to total energy expenditure (TEE). The doubly labelled water (DLW) technique is regarded as the gold standard for measuring TEE in a free-living context [10, 4, 1]. Total energy expenditure is composed of energy expended at rest, often referred to as basal energy expenditure (BEE), energy expended above the resting level due to PA, referred to as activity energy expenditure (AEE), and the thermic effect of food (TEF). The DLW technique requires the use of stable water isotopes and use of sophisticated laboratory equipment for estimating isotope enrichments over time in biological samples. The cost of dosing, sampling and laboratory analysis limit the use of the DLW technique in large epidemiological studies [4]. Consequently, use of affordable methods for assessing PA is becoming popular. A commonly used technique to assess PA objectively is accelerometry [4, 5]. An accelerometer is an electric motion sensor that monitors body acceleration due to PA [5]. It is, however, widely recognized that accelerometers underestimate some aspects of physical activities, such as swimming, cycling, sedentary activities and static exercise in free-living individuals [11, 4, 1, 12, 13]. Validation studies on PA, therefore, use the DLW technique to validate instruments for assessing usual activity in free-living individuals [1, 4]. The DuPLO Dutch study is one such validation study, where PA was assessed with a triaxial GT3X accelerometer (Actigraph, Pensacola, Florida) to monitor body acceleration in three axes [14], where DLW was used as a gold standard for measuring TEE. We used the DuPLO study to validate the measurement of TEE with the triaxial GT3X accelerometer.

The three main effects of the measurement error in TEE are: (i) bias in the mean level of TEE, (ii) loss of statistical power to detect a significant association between TEE and a health outcome, such as obesity [7], and (iii) bias in the association between TEE and a health outcome. The mean bias can be quantified with the mean discrepancy between the true and the measured activity level in the study population, and the loss of statistical power with the correlation coefficient between the measured and the true usual activity level (hereafter, validity coefficient); the bias in the association

can be quantified with the attenuation factor [8, 9]. A valid instrument will measure usual activity with a minimal bias, and a validity coefficient and attenuation factor close to one.

Currently, there is inadequate in-depth research on validation of the recently developed GT3X accelerometer for measuring TEE in free-living individuals. Moreover, current studies on other models of triaxial accelerometers stop at computing (i) the difference in the mean of accelerometer-derived measurements and DLW-derived measurements and (ii) the correlation coefficient between measurements from an accelerometer and DLW. Even worse, the error-prone accelerometers are often used to validate other instruments, such as physical activity questionnaires, which lead to erroneous validity measures and overestimation of validity of these instruments. For adequate validity assessment, however, a researcher needs to determine the magnitude of the validity coefficient associated with the use of the accelerometer to quantify loss of statistical power, and the attenuation factor to quantify the bias in the parameter estimate that quantifies the association [9, 8].

We assessed the validity of the accelerometer used in the DuPLO validation study as follows. First, we applied a plausible measurement error model to quantify the measurement error associated with the use of the accelerometer. Second, we estimated the bias in the mean level of TEE, the validity coefficient and the attenuation factor for the accelerometer-derived TEE. Third, we estimated these quantities conditional on subject characteristics, as this is the type of validity measurement that is relevant in epidemiological studies. Fourth, we proposed a calibration method for the accelerometer and estimated BEE to reduce loss of statistical power, attenuation and bias in the mean level of TEE caused by measurement error. Lastly, we assessed the performance of two prediction equations commonly used to calculate AEE from the accelerometer data.

MATERIAL AND METHODS

DuPLO STUDY

The DuPLO study participants consisted of a sub-sample from the NQplus study – a longitudinal study on diet and health (<https://www.wageningenur.nl/en/project/nqplus.htm>). The DuPLO study participants were recruited via email invitation, and were all Dutch, aged 20-70 years and living in Wageningen, Ede, Renkum and Arnhem [14]. The study was approved by the medical ethics committee of Wageningen University and Research. The purpose of the study was explained to the participants and written informed consent was obtained from each participant. Among the eligible participants, 200 agreed to participate in the DuPLO study (92 men, 108 women), out of which 154 agreed to participate in the accelerometer study and wore the accelerometer on the hip for 7 consecutive days. Out of the 200 DuPLO study participants, 70 agreed to participate in the DLW study, but due to physiologically implausible body water changes between repeated DLW measurements while the body weight remained stable, one participant was excluded from analysis. Thus, out of 69 participants, energy expenditure was measured by DLW (over 11 consecutive days) and 29 of these participants completed a second DLW measurement to estimate the within-individual variability. The participants joining the DLW study also wore the accelerometer, either during or after the DLW period (see Fig. 1). Data were collected from 2011 to 2013.

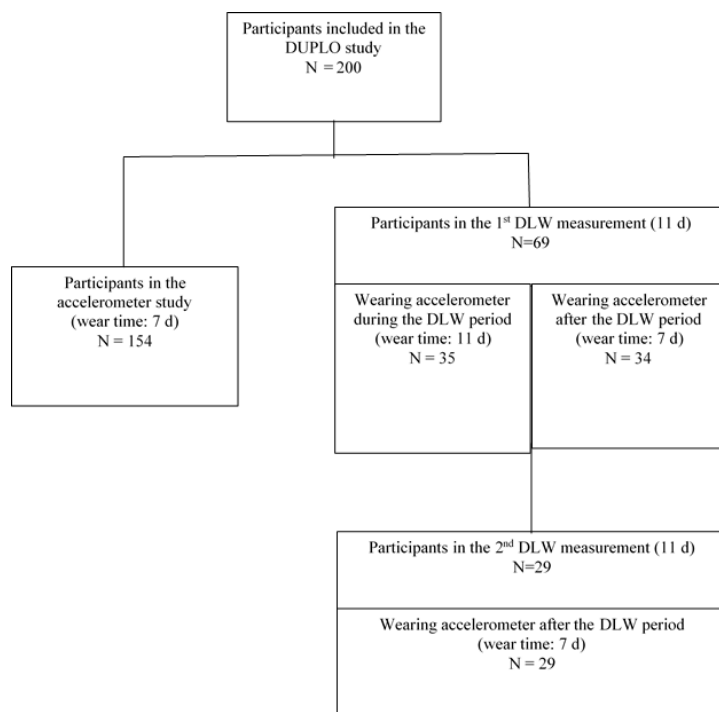


Fig. 1. The DuPLO study design, Netherlands, 2011-2013

ASSESSMENT OF TOTAL ENERGY EXPENDITURE WITH THE DLW

Doubly labelled water was used to measure TEE using a two-point protocol [15]. Participants were not eligible to join the DLW study if they were planning to travel abroad, were on an energy-restricted diet, used diuretics, lactated, were pregnant or planning to be pregnant during the study period, and if they were suffering from congestive heart failure, kidney failure or malabsorption. A day before the DLW dose, participants were instructed to follow a normal dietary pattern, refrain from alcohol, heavy exercise and exposure to high temperatures, and to stay in a fasting state the evening prior to DLW dosing. During the first visit, weight and height were measured and baseline urine and saliva samples were collected followed by ingestion of a dose of DLW. Participants received a mixture of 1.8 g 10% enriched H_2^{18}O (Centre for Molecular Research Ltd, Moscow, Russia) and 0.12 g 99.8% enriched $^2\text{H}_2\text{O}$ (Cambridge Isotope Laboratories, Inc, Andover, MA, USA) per kg body water. Body weights of male and female were assumed to contain 55% and 50% body water, respectively [16]. Additional urine and saliva samples were collected three and four hours post dose. Participants revisited the study center eleven days after dosing. At the second visit, body weight was re-measured, and two samples of urine and saliva were collected with one-hour interval between samples. To quantify within-individual variability in DLW measurements, second replicate measurements of DLW were taken from 30 participants (mean time between two measurements ~ 5 months). The samples were analyzed at the Centre for Isotope Research, Groningen, the Netherlands [17]. The rate of carbon dioxide production (rCO_2) was calculated as: $\text{rCO}_2 \text{ (L/day)} = (\text{TBW} / 2.078)(1.01 \text{ kO} - 1.04 \text{ kD}) - 0.0246 \text{ rGf}$, where TBW is total body water, kO and kD are isotope elimination rates of oxygen and deuterium, respectively, and $\text{rGf} = 1.05 \text{ TBW}(\text{kO} - \text{kD})$ [18]. Total energy expenditure from the DLW was calculated using the modified Weir equation: $\text{TEEdlw (kcal/day)} = \text{rCO}_2 \text{ (L/day)} \times (1.1 + 3.90/\text{RQ})$, where RQ was assumed to be 0.85 [19].

ASSESSMENT OF TOTAL ENERGY EXPENDITURE WITH AN ACCELEROMETER

A total of 154 individuals agreed to participate in the accelerometer study. For the DLW participants and the participants who solely joined the accelerometer study, a triaxial GT3X accelerometer was used to monitor PA. Participants received written instructions to wear the accelerometer on the hip for a minimum of 7 consecutive days. Accelerometers were not worn during sleeping and water activities. On wear time days, participants kept a diary about their daily activities and monitored non-wear time due to sleeping and water activities. Accelerometer raw data were analysed using ActiLife software version 6.6. Wear time was validated, where a non-wear period was defined as a minimum length of inactivity of 60 minutes. Daily AEE_{accel} was calculated from raw accelerometer activity data using two algorithms for AEE: (i) Freedson VM3 combination (2011) that uses activity data from all the three axes (hereafter, Freedson 2011) [20], and (ii) Freedson combination (1998) that uses activity data from one axis only (hereafter, Freedson 1998) [21]. The Freedson VM3 combination (2011) algorithm combines the Freedson VM3 (11) algorithm [22] with the work-energy algorithm to calculate AEE. The Freedson combination (1998) algorithm uses Work-Energy Theorem to calculate the caloric expenditure below 1951 counts and an algorithm developed by Freedson in 1997 to calculate the energy expenditure above 1952 counts [20]. The total energy expenditure from the accelerometer was estimated as $TEE_{accel} = (BEE + AEE_{accel})/0.9$ [23], where BEE was calculated from the participant's age, sex, height and weight using Henry's equation [24] and is hereafter referred to as BEE_{henry} ; the factor 0.9 follows from assuming a thermic effect of food as 10% of TEE [23, 4]. In the analysis, we excluded activity data for one participant who had implausible values.

MEASUREMENT ERROR MODEL FOR TOTAL ENERGY EXPENDITURE

We denote TEE measurements from the DLW for individual i on day j by R_{ij} , the corresponding accelerometer activity measurement by A_{ij} and a latent true usual activity for individual i by T_i . We relate A_{ij} and R_{ij} with T_i using a commonly used bivariate linear measurement error model [25] as

$$\begin{cases} A_{ij} = \beta_0 + \beta_A T_i + r_{Ai} + \varepsilon_{Aij} \\ R_{ij} = T_i + \varepsilon_{Rij} \end{cases}, \text{ where } \begin{cases} r_{Ai} \sim N(0, \sigma_{r_A}^2) \\ \varepsilon_{Aij} \sim N(0, \sigma_{\varepsilon_A}^2) \\ \varepsilon_{Rij} \sim N(0, \sigma_{\varepsilon_R}^2) \end{cases} \quad (1)$$

the intercept term β_0 reflects constant bias in the accelerometer that is independent of T_i and other terms in the model and is referred to as constant bias term; the slope β_A represents average population bias that is related with T_i , and is referred to as proportional scaling bias term; β_0 and β_A are jointly referred to as systematic bias terms [26]; r_{Ai} denotes random deviation of an individual's average bias relative to the average bias in the population and is referred to as person-specific bias [8, 27]; ε_{Aij} denotes within-individual random deviation from an individual's average bias; ε_{Rij} represents within-individual random deviation of DLW measurements from true level of usual activity. We further assumed independence between random terms, between each random error component and true usual activity, and between replicate measurements from the same instrument. True usual activity is distributed as $T_i \sim N(\mu_T, \sigma_T^2)$; A_{ij} is distributed as $A_{ij} \sim N(\beta_0 + \beta_A \mu_T, \beta_A^2 \sigma_T^2 + \sigma_{r_A}^2 + \sigma_{\varepsilon_A}^2)$ with

a mean that is biased for true mean μ_T . In contrast, for R_{ij} , we assumed no proportional scaling bias and no person-specific bias. Thus, measurement error in R_{ij} is purely due to within-individual random variation. In epidemiological studies, analyses on relations with PA are usually adjusted for confounding effects of individual characteristics, here, age, sex and the body mass index. In such analysis, the relevant validity measures are those depending on these characteristics. To calculate such conditional validity measures, we reparametrize the distribution of T_i as

$$T_i \sim N(\alpha_0 + \alpha_Z^T \mathbf{Z}, \sigma_T^2) \quad (2)$$

where $\alpha_0 + \alpha_Z^T \mathbf{Z} = \mu_T$, \mathbf{Z} is a vector of covariates consisting of individual characteristic variables with a vector of fixed effect parameters α_Z^T . Noteworthy, a more general measurement error model presented in (1) would include covariate effects in the A_{ij} component. We, however, did not include these covariates due to the complexity of the model and a difficulty with the model convergence given the relatively small sample size of the DuPLO study.

QUANTIFICATION OF THE MEASUREMENT ERROR IN THE ACCELEROMETER

The measurement error in accelerometer-derived TEE can be quantified in terms of the discrepancy between true mean TEE, as defined by the DLW (gold standard) method, and mean TEE as estimated from the accelerometer, i.e., with the bias. We explored the bias in mean TEE measurements based on the accelerometer as follows. First, for each subject with two replicate measurements from the accelerometer and the DLW, we plotted the mean TEE estimate from the accelerometer versus the mean TEE estimate from the DLW (hereafter, mean plot). Second, for each subject, we plotted the difference between mean TEE estimates from both instruments (as a measure of bias) versus the mean estimate from the DLW (as true TEE) in a Bland-Altman plot [28–30]. In the Bland-Altman plot, we computed 95% limits of agreement between the accelerometer-based and DLW measurements. The 95% limits of agreement, defined as mean difference ± 1.96 standard deviation of the difference, quantify the level of agreement between TEE as measured with both instruments. We further explored the structure of the measurement error in each instrument using Bland-Altman plots as explained further in the text. Noteworthy, to quantify the overall bias in the accelerometer-derived TEE, the constant bias and the systematic bias terms should be interpreted contemporaneously. From model (1), the overall mean bias can be estimated as

$$\widehat{\text{bias}} = \widehat{\beta}_0 + (\widehat{\beta}_A - 1)\widehat{\mu}_T \quad (3)$$

From equation (3), the overall bias = 0 when $\widehat{\mu}_T = \widehat{\beta}_0 / (1 - \widehat{\beta}_A)$

the overall bias is positive (overestimation of TEE) when $\widehat{\mu}_T = \widehat{\beta}_0 / (1 - \widehat{\beta}_A)$;

and the bias is negative (underestimation of TEE) when $\widehat{\mu}_T = \widehat{\beta}_0 / (1 - \widehat{\beta}_A)$

We proposed the following method to calibrate the accelerometer for measuring AEE and BEE_{henry} as estimated with Henry's equation. The reason for the calibration is to reduce the effect of error in accelerometer-based AEE and BEE_{henry}. We calibrated AEE derived from the accelerometer (AEE_{accel}, biased) and BEE_{henry} using TEE from the DLW (TEE_{dlw}, unbiased) by fitting the following regression calibration model:

$$E(0.9 \times \text{TEE}_{\text{dlw}} | \text{AEE}_{\text{accel}}, \text{BEE}_{\text{henry}}) = a_0 + a_1 \text{AEE}_{\text{accel}} + a_2 \text{BEE}_{\text{henry}}$$

and recalculate calibrated TEE from the accelerometer AEE and BEE_{henry} as

$$\text{TEE}_{\text{accel}} = (\hat{a}_0 + \hat{a}_1 \text{AEE}_{\text{accel}} + \hat{a}_2 \text{BEE}_{\text{henry}}) / 0.9$$

The loss of statistical power to detect a significant association between TEE and a health outcome due to the measurement error in TEE can be quantified with the validity coefficient [8]. The validity coefficient is a correlation coefficient between the measured and the true level of TEE, and can be expressed in terms of the measurement error model parameters as

$$\rho_{AT} = \frac{\text{cov}(A, T)}{\sqrt{\text{var}(T) \text{var}(A)}} = \frac{\beta_A \sigma_T}{\sqrt{\beta_A^2 \sigma_T^2 + \sigma_{rA}^2 + \sigma_{\varepsilon_A}^2}} \quad (4)$$

where ρ_{AT} is usually between zero and one; a value close to zero signifies a substantial loss in statistical power, meaning that the sample size required to detect a significant association will be inflated by a factor of $1/\rho_{AT}^2$.

The association between TEE and a health outcome might be biased, typically toward the null value when TEE is measured with error. The bias toward the null phenomenon is referred to as attenuation [31]. The magnitude of attenuation can be quantified with the attenuation factor, λ_A . When the relation between measured and true usual TEE is linear, λ_A is the regression slope of true on measured TEE and is expressed in terms of model parameters as

$$\lambda_A = \frac{\text{cov}(A, T)}{\text{var}(A)} = \frac{\beta_A \sigma_T^2}{\beta_A^2 \sigma_T^2 + \sigma_{rA}^2 + \sigma_{\varepsilon_A}^2}, \quad (5)$$

where a λ_A value close to zero indicates severe attenuation, meaning that the estimated parameter that quantifies the association will be smaller than the true value that would be observed if TEE were measured without error. To adjust

DESCRIPTIVE STATISTICAL ANALYSES AND MODEL FITTING

We summarized mean TEE data from the DuPLO study with a mixed model approach to handle the imbalance in the study design, because some participants did not have complete measurements. We estimated error distributions by computing within-individual differences for TEE data derived from the DLW and the accelerometer separately. We explored error distributions with histograms, kernel density plots, Bland-Altman plots, and formally with the Shapiro-Wilk test. The bias in the accelerometer-based TEE was explored with the mean plot and the Bland-Altman plot. We subsequently fitted the proposed bivariate

measurement error model using a maximum likelihood method with the Newton-Raphson optimization technique and the adaptive Gaussian quadrature with 10 quadrature points. Note that in the bivariate model, we jointly fitted the accelerometer and DLW data simultaneously. The method was implemented in SAS version 9.3 using the NLMIXED procedure.

RESULTS

Presented in Table 1 are summary measures for the DuPLO main study ($n = 200$), and the DLW validation study variables ($n = 69$). On average, participants in the validation study had similar characteristics (age, height, weight and BMI) as participants in the main study. Additionally, male participants in the main study had higher mean age, body weight, height and body mass index than their female counterparts.

Table 1. Overall and sex-specific mean (standard deviation) and number of observations (n) for subject characteristics in the DuPLO main study and in the DLW validation study, Netherlands, 2011-2013

Variables	Study	Overall ($n^a = 200$; $n^b = 69$) mean (SD)	Male ($n^a = 92$; $n^b = 37$) mean (SD)	Female ($n^a = 108$; $n^b = 32$) mean (SD)
Age, years	DuPLO study ^a	55.7 (10.5)	58.2 (9.3)	53.6 (11.0)
	DLW validation ^b	57.1 (9.2)	58.2 (8.2)	55.9 (10.1)
Weight, kg	DuPLO study	76.0 (14.2)	83.1 (12.9)	69.9 (12.3)
	DLW validation	77.4 (13.7)	82.9 (13.2)	71.2 (11.7)
Height, m	DuPLO study	1.73 (0.08)	1.79 (0.06)	1.68 (0.06)
	DLW validation	1.74 (0.08)	1.79 (0.06)	1.69 (0.07)
BMI, kg/m ²	DuPLO study	25.2 (0.04)	25.8 (3.58)	24.7 (4.06)
	DLW validation	25.5 (3.57)	25.9 (3.44)	25.0 (3.70)

n^a DuPLO main study size; n^b DuPLO DLW validation study size

In Table 2, the mean and standard deviation for BEE, TEE and AEE are presented. The mean BEE estimate for male participants is greater than that for females, and the mean TEE derived from DLW is greater than the mean TEE derived from the accelerometer. For instance, the underestimation of mean TEE with the accelerometer ranges from 18% (using Freedson 2011 equation) to 20% (using Freedson 1998 equation) as compared with the DLW. Moreover, TEE and AEE derived from the accelerometer using Freedson 2011 equation (that measures body acceleration in three axes) are greater, on average, and with greater variability than those obtained using Freedson 1998 equation (that measures body acceleration in one axis).

Figure 2 displays Bland-Altman plots for TEE derived from (a) the accelerometer, (b) the DLW and from (d) both instruments, also shown is (c) the scatterplot of the mean TEE estimate from the accelerometer versus the mean estimate from the DLW. Note that in the figure, we present the accelerometer TEE data derived using Freedson 2011 equation to estimate AEE; a similar trend is shown using Freedson 1998 equation.

Table 2. Mean (standard deviation) for BEE, and predicted mean (standard deviation) for TEE from linear mixed models. DuPLO validation study, Netherlands, 2011-2013

	Overall mean (SD)	Male mean (SD)	Female mean (SD)
BEE ^a , kcal/day (n=153)	1508.4 (245.9)	1700.1 (200.5)	1345.1 (140.5)
TEE _{dlw} , kcal/day	2678.6 (343.2)	3047.6 (323.1)	2364.3 (360.3)
TEE _{accel} ^b , kcal/day	2185.1 (192.8)	2423.0 (179.4)	1982.5 (204.2)
TEE _{accel} ^c , kcal/day	2141.7 (186.9)	2389.5 (173.9)	1930.7 (197.9)
AEE _{accel} , Freedson 2011 ^b , kcal/day	453.9 (124.1)	486.0 (115.5)	426.6 (131.4)
AEE _{accel} , Freedson 1998 ^c , kcal/day	414.6 (114.7)	455.5 (106.7)	379.8 (121.5)

Abbreviation: BEE, basal energy expenditure; DLW, doubly labelled water; accel, accelerometer; AEE, activity energy expenditure; TEE, total energy expenditure;

^aBEE predicted from age, sex and weight using Henry's equation,

^bAccelerometer-derived TEE and AEE, where AEE is calculated from accelerometer data using Freedson VM3 (2011) combination equation;

^cAccelerometer-derived TEE and AEE, where AEE is calculated from accelerometer data using Freedson Combination (1998) equation.

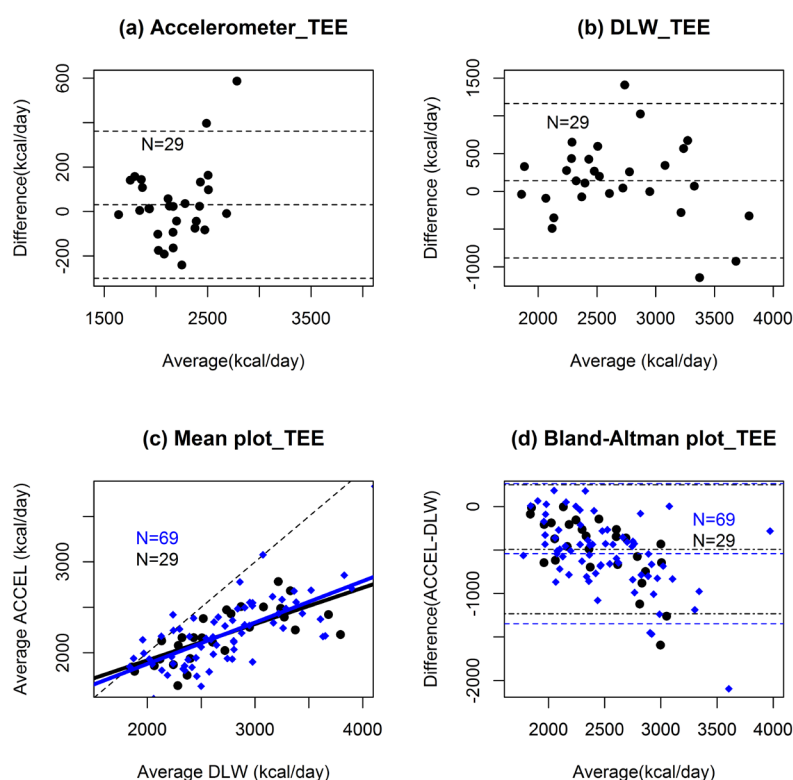


Fig. 2. Bland-Altman plots for total energy expenditure (TEE) measurements derived from (a) the accelerometer and (b) DLW, where within-subject differences are plotted against subject averages; also shown is the mean difference (middle dotted line) and 95% limits of agreement (extreme dotted lines); in (c), average TEE estimate from accelerometer for a subject is plotted against average TEE estimate from DLW; in (d) a subject's difference in the average TEE estimates from accelerometer and DLW is plotted against the corresponding subject average TEE estimate from DLW; the blue dots (N = 69) refers to first replicate measurements. DuPLO study, Wageningen, Netherlands, 2011-2013.

In Figures 2 (a) and (b), the scatter plots appear to be spread randomly and do not show any discernible trend. The lack of trend in the scatter plots suggests that the magnitude of random errors in the accelerometer and the DLW do not depend on the mean level of TEE. The fitted regression line shown in Figure 2 (c) suggests that TEE for participants with large mean DLW values are underestimated more with the accelerometer than for participants with small mean DLW values. The flattened regression slope further suggests the existence of a proportional scaling bias in the accelerometer for measuring TEE. The accelerometer-based method underestimates mean TEE by 492 kcal/day (Figure 2 (d), dotted middle line).

Based on these exploratory findings, we assumed normality and additivity for the distribution of within-individual errors in the TEE measurements from both accelerometer and DLW, and assumed a systematic bias in the accelerometer. We subsequently fitted the proposed bivariate measurement error model; first, by letting the true TEE activity depend on the subject's age, sex and BMI and, second, without conditioning on these subject characteristic variables. Note that age and BMI were standardized to improve the convergence of the model.

The parameter estimates from the measurement error model are presented in Table 3. There seems to be a constant bias in the accelerometer that is independent of an individual's level of activity and other terms in the model, though not statistically significant. For instance, when T_i is predicted conditional on the covariates, and the accelerometer-derived AEE is calculated using Freedson 2011 equation, TEE will be estimated with a constant bias of 149 kcal/day. From this analysis, the constant bias in the accelerometer ranges from 53kcal/day to 359 kcal/day. The estimate of β_A , the proportional scaling bias term that depends on an individual's level of TEE, is less than one for the covariate-adjusted analyses; this means that TEE for an individual with high energy expenditure will be underestimated more with the accelerometer than that of an individual with less energy expenditure. For instance, accelerometer-derived TEE that is predicted by the Freedson 2011 equation will underestimate the regression coefficient with the true TEE by about 34% ($\beta_A = 0.758$). From this analysis, the proportional scaling bias associated with the use of the accelerometer ranges from 33% ($\beta_A=0.664$) to 22% ($\beta_A=0.778$).

The results shown in Table 3 indicate that the accelerometer often underestimates the TEE. For instance, using the third line in Table 3, $\hat{\beta}_0/(1 - \hat{\beta}_A) = 387 / (1-0.67) = 1173$, which is less than the mean TEE from the DLW (see Table 2). This means that, in practice, for most individuals the TEE is underestimated when measured with an accelerometer.

Further shown in Table 3 are the person-specific bias σ_{r_A} and the within-person random error σ_{ε_A} in the accelerometer measurements. From the results presented in Table 4, it is evident that the accelerometer underestimates the mean TEE. For instance, the accelerometer underestimates mean TEE by about 500 kcal/day, when a model conditional on the covariates and using the Freedson 2011 prediction equation is used, corroborating the exploratory findings presented in Table 2 and Figure 2.

Table 3. Estimates for parameters of measurement error in the accelerometer-derived TEE and variance of true TEE is predicted with and without conditioning on subject's age, sex and BMI^a, where AEE from accelerometer is predicted with Freedson VM3 (2011) combination equation^b and Freedson Combination (1998) equation^c, DuPLO study, Wageningen, Netherlands, 2011-2013

True TEE conditional on the covariates ^a	Intercept of regression of accelerometer-based TEE on true TEE β_0 (SE)	Regression slope of regression of accelerometer-based TEE on true TEE β_A (SE)	Standard deviation of true TEE level σ_T (SE)	Standard deviation of person-specific bias in accelerometer-based TEE σ_{r_A} (SE)	Standard deviation of person-specific bias in accelerometer-based TEE σ_{e_A} (SE)	Standard deviation of within-person random error in DLW-based TEE σ_{e_R} (SE)
Yes ^b	149 (240)	0.758 (0.087)	228 (33)	88 (40)	109 (9)	405 (33)
Yes ^c	53 (246)	0.778 (0.089)	212 (32)	73 (46)	107 (9)	412 (33)
No ^b	387 (287)	0.670 (0.105)	480 (58)	151 (53)	110 (9)	371 (45)
No ^c	359 (279)	0.664 (0.102)	482 (58)	155 (51)	108 (9)	368 (44)

^a Whether a subject's true activity is predicted by conditioning on age, sex and BMI; TEE, total energy expenditure;
^b AEE used to estimate TEE is calculated using Freedson VM3 (2011) combination equation that uses accelerometer activity data from the three axes; ^c AEE used to estimate TEE is calculated using Freedson Combination (1998) equation that uses accelerometer activity data from one axis only.

Based on the validity coefficient estimates presented in Table 4, there will be a minimal loss of statistical power to detect a significant association between TEE and a health outcome, when accelerometer-derived TEE is used. For instance, to attain the required power to detect a significant association when ρ_{AT} is 0.78, the sample size of accelerometer study should be about 1.6 times as large as (i.e., $1/\rho_{AT}^2 = 1/0.78^2$) the sample size that would be required if TEE were measured exactly.

Table 4. Estimates for validity coefficient, attenuation factor and mean bias for accelerometer-derived TEE, when true TEE is predicted with and without conditioning on a subject's age, sex and BMI^a, where AEE from accelerometer is predicted with Freedson VM3 (2011) combination equation^b and Freedson Combination (1998) equation^c, DuPLO study, Wageningen, Netherlands, 2011-2013

True TEE conditional on the covariates	Correlation between accelerometer-based TEE and true TEE ρ_{AT} (SE)	Attenuation factor λ_A (SE)	Mean bias in accelerometer-based TEE measurements in kcal/day
Yes ^b	0.78 (0.089)	0.80 (0.177)	-502 (96)
Yes ^c	0.79 (0.094)	0.80 (0.186)	-545 (88)
No ^b	0.86 (0.068)	1.12 (0.133)	-513 (47)
No ^c	0.86 (0.067)	1.12 (0.133)	-558 (47)

^a Whether a subject's true activity is predicted by conditioning on age, sex and BMI; TEE, total energy expenditure;
^b AEE used to estimate TEE is calculated using Freedson VM3 (2011) combination equation that uses accelerometer activity data from the three axes; ^c AEE used to estimate TEE is calculated using Freedson Combination (1998) equation that uses accelerometer activity data from one axis only.

Similarly, when the fitted model is adjusted for the covariates, there will be some attenuation in the TEE-outcome associations that is associated with the use of the accelerometer. For instance, when TEE measurements based on the accelerometer with a λ_A value of 0.80 are used, a true relative risk of 0.6 would be observed as 0.66 ($0.6^{0.80}$). Conditioning on the covariates affects the magnitude and level of precision of parameter estimates from the measurement error model. Contrastingly, without adjusting for the covariates, λ_A will be overestimated by 12%, meaning that the observed association would appear stronger than the true association; this accentuates the need to adjust

correctly for the confounding effects of the covariates in measurement error modeling. Again, the mean bias in accelerometer-derived TEE using Freedson 1998 equation is greater than the mean bias using Freedson 2011 equation.

Presented in Table 5 are the results before and after calibrating the accelerometer-derived AEE and the estimated BEE and then recalculating TEE. The estimated calibration coefficients are:

$$\hat{a}_0 = 19.42, \hat{a}_1 = 0.497 \text{ and } \hat{a}_2 = 1.458.$$

With the proposed calibration method, the validity coefficient estimate increased from 0.78 to 0.95 and attenuation factor from 0.80 to 0.94; the absolute bias in the mean TEE is reduced from 502 kcal/day to 6 kcal/day. Note that these calculations yield too optimistic results, as they are calculated on the same data that were used to estimate the calibration coefficients.

Table 5. Parameter estimates for the measurement error model, validity coefficient, and attenuation factor and mean bias estimates for accelerometer-derived TEE using Freedson VM3 (2011) combination equation, with and without calibrating the accelerometer activity data, DuPLO study, Wageningen, Netherlands, 2011-2013

Accelerometer AEE calibrated ^a	Correlation between accelerometer-based TEE and true TEE ρ_{AT} (SE)	Attenuation factor λ_A (SE)	Mean bias in accelerometer-based TEE measurements in kcal/day
Not calibrated	0.78 (0.089)	0.80 (0.177)	-502 (96)
Calibrated	0.95 (0.014)	0.94 (0.107)	6 (18)

^a Whether a subject's true activity is predicted by conditioning on age, sex and BMI; TEE, total energy expenditure.

DISCUSSION

We validated the triaxial accelerometer (GT3X) in the DuPLO study against the DLW by calculating the bias in the mean TEE, the correlation coefficient between measured and true TEE and the magnitude of attenuation in the association between physical activity as measured by TEE and a health outcome of interest. The accelerometer underestimated TEE by about 20% on average as compared with the DLW, which is within the 95% confidence interval reported in a review study by Van Remoortel et al. [11] and consistent with the findings from other similar studies [5, 1]. The magnitude of underestimation of TEE in free-living individuals when AEE is calculated using Freedson 1998 is more than the mean estimate using Freedson 2011 equation. The observed underestimation with the accelerometer relative to the DLW could be due to too simplistic prediction equations for accelerometer-derived AEE [5], and low sensitivity of the accelerometer to monitor sedentary activities, bicycling and static exercise, especially when worn at the waist [4, 32]. The DuPLO study participants bicycled regularly and failure to monitor bicycling could have resulted in more underestimation than in studies conducted in other countries. It is noted that without monitoring fidgeting alone, an individual's daily TEE could be underestimated by up to 800 kcal/day [13].

The DuPLO study analysis revealed that accelerometer underestimated true mean TEE, especially for physically very active individuals. From the covariate-adjusted analysis, we found even lower attenuation coefficients. Previous studies showed similar findings when physical activity was assessed with questionnaires [8]. This, therefore, shows the importance of covariate-adjustment in validating PA instruments.

Presently, the study by Ferrari et al. [8] is the closest to our study. It is, however, difficult to compare our results quantitatively with those of the mentioned study, because the authors expressed physical activity in log-transformed MET hours per week as opposed to untransformed kcal/day, and they assumed PA logs as the reference measure as opposed to DLW in our study.

In our analysis, conditioning on the subject's characteristics influences the validity measures, in line with findings from the literature [8]. Subject's sex, age and body mass index explained part of the between-individual variability in the activity level. With higher variability in true values, the correlation between true and measured TEE increases. In Table 4, for example, the validity coefficient for the model without covariates is larger than that of the model with the covariates. This is because covariates explain part of the correlation. If the model is conditioned on the covariates, the validity coefficient represents the partial correlation coefficient between true and estimated TEE given the covariates. In epidemiologic analysis, it is common to adjust for the confounding effects of these characteristic variables, and therefore the validation coefficient conditional on confounding variables is the validation coefficient of interest.

The magnitudes of the validity coefficient and the attenuation factor estimates seemed similar, irrespective of the prediction equation used to calculate AEE from the accelerometer data. The similarity in the estimates suggests minimal contribution of activity data recorded on all three axes over the activity data recorded on one axis. This finding is in line with previous studies that showed minimal improvement when AEE was measured with a triaxial accelerometer over a uniaxial one [4], or by using one prediction equation over the other [13, 33].

This study provides an in-depth description of measurement error in the accelerometer activity data and essential components of plausible measurement error structure in validating the GT3X accelerometer model. The proposed calibration approach is intuitive and corrects for the measurement error in the accelerometer measurements and estimated BEE in the DuPLO study population. However, whether this applies to other populations needs further investigation.

This study had a few limitations. First, its external validity is limited, because DuPLO participants were of similar ethnicity, living in the same region and were all adults. Thus, generalizing the study findings to different populations might be misleading. Second, BEE was predicted with an equation, which can result in an additional error whose effect requires in-depth scrutiny. There are more reliable but expensive methods to measure BEE such as indirect calorimetry in a controlled environment. Lastly, the algorithms used here to calculate AEE from the accelerometer raw data are too simplistic and might not capture all aspects of physical activity energy expenditure as measured by the accelerometer; therefore, more sophisticated methods, such as neural networks, should be developed to estimate AEE from the triaxial accelerometer data [22].

CONCLUSION

The accelerometer underestimated mean TEE as compared with the DLW in the DuPLO study population. Given the measurement error model used in this study, there would be minimal loss in statistical power and bias in the association between TEE and a health outcome, when TEE is assessed with the GT3X accelerometer.

REFERENCES

- [1] Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity*. 2007;15(10):2371-9. doi: 10.1038/Oby.2007.281.
- [2] WHO. Physical Activity. WHO. 2015. http://www.who.int/topics/physical_activity/en/. Accessed March, 27 2015.
- [3] Leinonen AM, Ahola R, Kulmala J, et al. Measuring physical activity in free-living conditions - Comparison of three accelerometry-based methods. *Front Physiol*. 2017;7. doi:10.3389/fphys.2016.00681.
- [4] Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: An overview of objective measures. *Front Nutr*. 2014;1:5.
- [5] Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: New insights and validation studies. *Obes Rev*. 2013;14(6):451-62. doi:Doi 10.1111/Obr.12021.
- [6] Nusser SM, Beyler NK, Welk GJ, Carriquiry AL, Fuller WA, King BMN. Modeling errors in physical activity recall data. *J Phys Act Health*. 2012;9:S56-S67.
- [7] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models. New York: Chapman& Hall/CRC; 2006.
- [8] Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol*. 2007;166(7):832-40. doi:Doi 10.1093/Aje/Kwm148.
- [9] Tooze JA, Troiano RP, Carroll RJ, Moshfegh AJ, Freedman LS. A Measurement error model for physical activity level as measured by a questionnaire with application to the 19992006 NHANES Questionnaire. *Am J Epidemiol*. 2013;177(11):1199-208. doi:Doi 10.1093/Aje/Kws379.
- [10] Hallal PC, Reichert FF, Clark VL, et al. Energy expenditure compared to physical activity measured by accelerometry and self-report in adolescents: A validation study. *PLoS One*. 2013;8(11). doi:ARTN e77036; DOI 10.1371/journal.pone.0077036.
- [11] Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act*. 2012;9:84. doi:10.1186/1479-5868-9-84.
- [12] Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol*. 2011;111(2):187-201. doi:DOI 10.1007/s00421-010-1639-8.
- [13] Leenders NYJM, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. *Medicine and Science in Sports and Exercise*. 2001;33(7):1233-40.
- [14] Trijsburg L, de Vries JHM, Boshuizen HC, et al. Comparison of duplicate portion and 24 h recall as reference methods for validating a FFQ using urinary markers as the estimate of true intake. *Brit J Nutr*. 2015;114(8):1304-12. doi:10.1017/s0007114515002871.
- [15] IAEA. Assessment of Body Composition and Total Energy Expenditure in Humans Using Stable Isotope Techniques. AEA Human Health Series No. 3. Vienna: International Atomic Energy Agency 2009.
- [16] Chumlea WC, Guo SS, Kuczmarski RJ, et al. Body composition estimates from NHANES III bioelectrical impedance data. *Int J Obesity*. 2002;26(12):1596-609. doi:DOI 10.1038/sj.ijo.0802167.
- [17] Guidotti S, Jansen HG, Aerts-Bijma AT, Verstappen-Dumoulin BMAA, Van Dijk G, Meijer HAJ. Doubly Labelled Water analysis: Preparation, memory correction, calibration and quality assurance for ^{62}H and ^{618}O measurements over four orders of magnitudes. *Rapid Communications in Mass Spectrometry*. 2013;27(9):1055-66.
- [18] Schoeller DA, Leitch CA, Brown C. Doubly Labeled Water Method - Invivo Oxygen and Hydrogen Isotope Fractionation. *Am J Physiol*. 1986;251(6):R1137-R43.
- [19] Weir JBD. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol-London*. 1949;109(1-2):1-9.
- [20] What is the difference among the Energy Expenditure Algorithms?. <https://actigraph.desk.com/customer/en/portal/articles/2515835-what-is-the-difference-among-the-energy-expenditure-algorithms->. Accessed November 8 2017.
- [21] Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, Inc. Accelerometer. *Med Sci Sport Exerc*. 1998;30(5):777-81. doi:10.1097/00005768-199805000-00021.
- [22] Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14(5):411-6. doi:DOI 10.1016/j.jsams.2011.04.003.
- [23] Neuhauser ML, Di C, Tinker LF, et al. Physical activity assessment: Biomarkers and self-report of activity-related energy expenditure in the WHI. *Am J Epidemiol*. 2013;177(6):576-85. doi:10.1093/aje/kws269.
- [24] Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. *Public Health Nutr*. 2005;8(7A):1133-52.

- [25] Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol.* 2003;158(1):14-21.
- [26] Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol.* 2001;153(4):394-403. doi:DOI 10.1093/aje/153.4.394.
- [27] Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol.* 2003;158(1):14-21; discussion 2-6.
- [28] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
- [29] Lim S, Wyker B, Bartley K, Eisenhower D. Measurement error of self-reported physical activity levels in new york city: assessment and correction. *Am J Epidemiol.* 2015;181(9):648-55. doi:10.1093/aje/kwu470.
- [30] Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. *Statistic Med.* 2008;27(5):778-80. doi:10.1002/sim.3086.
- [31] Kipnis V, Freedman L, Brown C, Hartman A, Schatzkin A. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol.* 1997;146(10):842-54.
- [32] Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sport Exerc.* 2000;32(9):S450-S6. doi:10.1097/00005768-200009001-00003.
- [33] Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol.* 2006;98(6):601-12. doi:10.1007/s00421-006-0307-5.

Cite this article as:
 Agogo GO, van der Voet H, Hulshof PJM, van 't Veer P, Trijsburg L, van Eeuwig FA, Boshuizen HC.
 Validation of accelerometer for measuring physical activity in free
 living individuals. *Balt J Health Phys Act.* 2018;10(1):7-21.
 doi: 10.29359/BJHPA.10.1.01