

# Sleep Estimates Using Microelectromechanical Systems (MEMS)

Bart H. W. te Lindert, MSc<sup>1</sup>; Eus J. W. Van Someren, PhD<sup>1,2</sup>

<sup>1</sup>Department of Sleep and Cognition, Netherlands Institute for Neuroscience, an institute of the Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands; <sup>2</sup>Departments of Integrative Neurophysiology and Medical Psychology, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University and Medical Center, Amsterdam, The Netherlands

**Study Objectives:** Although currently more affordable than polysomnography, actigraphic sleep estimates have disadvantages. Brand-specific differences in data reduction impede pooling of data in large-scale cohorts and may not fully exploit movement information. Sleep estimate reliability might improve by advanced analyses of three-axial, linear accelerometry data sampled at a high rate, which is now feasible using microelectromechanical systems (MEMS). However, it might take some time before these analyses become available. To provide ongoing studies with backward compatibility while already switching from actigraphy to MEMS accelerometry, we designed and validated a method to transform accelerometry data into the traditional actigraphic movement counts, thus allowing for the use of validated algorithms to estimate sleep parameters.

**Design:** Simultaneous actigraphy and MEMS-accelerometry recording.

**Setting:** Home, unrestrained.

**Participants:** Fifteen healthy adults (23-36 y, 10 males, 5 females).

**Interventions:** None.

**Measurements:** Actigraphic movement counts/15-sec and 50-Hz digitized MEMS-accelerometry.

**Analyses:** Passing-Bablok regression optimized transformation of MEMS-accelerometry signals to movement counts. Kappa statistics calculated agreement between individual epochs scored as wake or sleep. Bland-Altman plots evaluated reliability of common sleep variables both between and within actigraphs and MEMS-accelerometers.

**Results:** Agreement between epochs was almost perfect at the low, medium, and high threshold ( $\kappa = 0.87 \pm 0.05$ ,  $0.85 \pm 0.06$ , and  $0.83 \pm 0.07$ ). Sleep parameter agreement was better between two MEMS-accelerometers or a MEMS-accelerometer and an actigraph than between two actigraphs.

**Conclusions:** The algorithm allows for continuity of outcome parameters in ongoing actigraphy studies that consider switching to MEMS-accelerometers. Its implementation makes backward compatibility feasible, while collecting raw data that, in time, could provide better sleep estimates and promote cross-study data pooling.

**Keywords:** accelerometry, actigraphy, cohort studies, sleep

**Citation:** te Lindert BHW; Van Someren EJW. Sleep estimates using microelectromechanical systems (MEMS). *SLEEP* 2013;36(5):781-789.

## INTRODUCTION

The detailed analysis of human sleep requires polysomnography (PSG), the simultaneous recording of the electroencephalogram (EEG), electrooculogram, electromyogram and often other physiologic signals. Although PSG remains the gold standard, the procedure is time-consuming and costly. This unfortunately makes it difficult to use PSG to quantify sleep in the large-scale cohorts that are, for example, desirable in twin-sibling heritability studies and genome-wide association studies (GWAS).

A step up in cost-efficacy, yet a step down in precision, was realized with the introduction of actigraphy, the recording of wrist movements with a small solid-state recorder.<sup>1-3</sup> Given the association between prolonged immobility and sleep,<sup>4</sup> actigraphy is considered a reasonably reliable and valid alternative method to estimate sleep-wake patterns.<sup>5</sup> In the days that actigraphy was developed, miniature solid-state memory for data storage was of very limited capacity. Therefore, the detected acceleration signal is usually preprocessed to store a single count value per

epoch, usually per 30 sec or per min. Moreover, linear three-dimensional accelerometers were quite bulky and costly, so that manufacturers usually resorted to monoaxial pressure sensitive piezo elements, e.g., with a lever connected to it, yielding a nonlinear acceleration response and some between-device variability. Still, acceptably valid sleep estimates could be derived by algorithms acting on the recorded counts,<sup>2,6,7</sup> although it has been argued whether sensitivity and specificity are acceptable in people with disorders such as insomnia<sup>8-10</sup> or Parkinson disease.<sup>11-13</sup> Actigraphy does not provide the detailed information on sleep that PSG gives, but has a number of advantages over PSG. First, the unobtrusive character of actigraphy makes it tolerable to uncooperative patients.<sup>14-16</sup> Second, the possibility to record for multiple days or even weeks has opened up the possibility to quantify between-day sleep variability patterns and to obtain more representative sleep estimates than can be obtained with one or at most only a few nights of PSG.<sup>17,18</sup> Unfortunately, the pooling and integration of actigraphic data from several cohorts, as is for example considered necessary for GWAS studies where actigraphy can be of great value,<sup>19</sup> is obstructed by brand-specific differences in instrumentation, data reduction, and scoring algorithms.<sup>20</sup>

These limitations can be overcome with the help of recent advances in microelectronics. Accelerometers integrated in microelectromechanical systems (MEMS) have become widely available, thanks to their mass production for use in mobile devices such as smartphones, games, and tablet computers. This provides the opportunity for long-term recording of raw triaxial

Submitted for publication June, 2012

Submitted in final revised form December, 2012

Accepted for publication December, 2012

Address correspondence to: Bart H.W. te Lindert, MSc, Netherlands Institute for Neuroscience (NIN), Meibergdreef 47, 1105 BA, Amsterdam; Tel: +31(0) 20-5665500; Fax: +31 (0)20-5666121; E-mail: b.te.lindert@nin.knaw.nl

linear accelerometry signals at high sample rates in SI units, rather than brand-specific actigraphy count signals extracted from nonlinear uniaxial accelerometers, opening up the possibility to compare and pool data from any brand. Although accelerometers with good linearity have been applied in actigraphs,<sup>21</sup> the long-term high-resolution sampling of three-dimensional linear accelerometers has only become widely available with MEMS. Importantly, as is the case for EEG, reproducible algorithms for preprocessing and analysis and later reanalysis of raw signals can be published and used.

MEMS-accelerometers for long-term recordings have become available and shown valid to estimate energy expenditure.<sup>22</sup> Currently, several larger cohort studies and follow-up studies use the traditional activity count type of actigraphs. The necessity of continuity of outcome measures may impede researchers from switching to the newer type and attain other advantages of the more precise movement assessment. The continuity problem would be solved with a valid method to convert the MEMS-accelerometry signal into the traditional counts. This would subsequently allow for the use of the very same validated algorithms to estimate sleep parameters from activity counts and thus provide backward compatibility in long-term follow-up studies and cohort studies.

Therefore, the aim of the current study is to provide a data-driven optimized conversion of the raw accelerometry signal into counts as stored in one of the most commonly used actigraphs, the Actiwatch (Cambridge Neurotechnology Ltd., Cambridge, UK and Mini Mitter, Respironics Inc., Bend, Oregon, USA). Subsequently, the performance of the generated counts in estimating sleep parameters will be evaluated by comparing them with sleep parameter estimates obtained from the simultaneously worn actigraphs. Finally, the reliability of MEMS-accelerometry recorders as compared to traditional actigraphs will be evaluated by comparing sleep estimate agreement between two MEMS-accelerometers, two actigraphs, and a MEMS-accelerometer and an actigraph.

## METHODS

### Study Participants

Fifteen volunteers (10 males,  $29.7 \pm 3.9$  y [23-36 y], mean  $\pm$  SD [range]) were recruited by advertisement and word of mouth. All participants stated to be in good health and worked regular office hours. As part of a larger protocol, the study was approved by the Ethics Committee of the VU University and Medical Center.

### Actigraphs and MEMS-Accelerometers

Traditional actigraphic count recordings were obtained using the Actiwatch. MEMS-accelerometry signals were obtained using the Geneactiv recorder (ActivInsights Ltd., Kimbolton, UK).

The Actiwatch contains an acceleration-responsive piezoelectric sensor with a range of  $\pm 5$  g and a sensitivity  $\geq 0.01$  g. Sensitivity to motion is greatest when the wrist is moved along the palmar-dorsal axis, less in the radial-ulnar axis, and least when the wrist moves parallel to the long axis of the radius and ulna. This generates an intrinsic nonlinearity, which is further amplified by the use of a lever to amplify the force on the piezoelectric sensor. The Actiwatch applies a 3-11 Hz analog band-

pass filter prior to digital sampling at 32 Hz and a resolution of approximately 25 counts/g. The signal is converted online to a summary count measure by taking the peak value of each sec and determining the sum across the epoch length. The current investigation used 15-sec epochs, the best available time resolution on the Actiwatch.

The Geneactiv recorder contains a triaxial MEMS-accelerometer with a range of  $\pm 8$  g and a sensitivity of  $\geq 0.004$  g. It records both motion-related and gravitational acceleration and has a linear and equal sensitivity along the three axes. The x-axis of the Geneactiv recorder corresponds to the radial-ulnar axis, the y-axis to the long axis of the radius and ulna, and the z-axis to the palmar-dorsal axis. Therefore, the z-axis of the Geneactiv recorder corresponds to the most sensitive axis of the Actiwatch. The sampling frequency of the Geneactiv recorder was set at 50 Hz.

### Procedures

Each study participant underwent an overnight home recording from 18:00 to 09:00 the next day. Each study participant kept a sleep log of their sleep-wake schedule (including bedtime, lights out time, final wake time, and get-up time). No instructions were given with regard to their sleep pattern. All recordings were obtained within 5 weeks (week 29-34, 2011). For each recording, participants wore two Geneactiv recorders and two Actiwatches simultaneously on the nondominant wrist. Prior to each recording, two Geneactiv recorders and two Actiwatches were randomly selected from a pool of six Geneactiv recorders and a pool of six Actiwatches. To optimize synchronization, the same personal computer was used to initiate all devices sequentially. Synchronization was verified by cross-correlation analysis of the time series and corrected using the optimal lag if required. The two Geneactiv recorders and two Actiwatches were tightly attached to a 28 mm  $\times$  69 mm  $\times$  1 mm rigid plastic strip (Figure 1). The strip was oriented parallel to the long axis of the forearm. Due to the shape of the devices, both Actiwatches were attached to the medial (ulnar) side of the strip, whereas the Geneactiv recorders were placed on the lateral (radial) side. The devices lay directly on the dorsal surface of the wrist underneath the strip. The device pairs were placed side by side on the wrist, and randomly assigned to the more proximal and more distal location. Thus, a distal pair of one Actiwatch and one Geneactiv recorder was placed as close as possible to the hand, and the other pair was placed immediately proximal to that.

### Data Conversion Steps

As a first step to obtain a transfer function that converts raw Geneactiv accelerometry data into Actiwatch counts, accelerometry data were preprocessed according to the technical specification of the Actiwatch. Thus, only the palmar-dorsal (z) axis of the Geneactiv recorder was used for analysis as it corresponds to the most sensitive axis of the Actiwatch. Subsequently, a band-pass filter (3-11 Hz, Butterworth, order 5) was applied, which removed most of the acceleration signal caused by rotations in the gravitational field.<sup>23</sup> The signal was then rectified and divided into 128 bins between 0 and 5 g, such that a value of 25 corresponds to approximately 1 g. Within each sec the peak values were detected and summed to obtain one count value per 15-sec epoch.

An initial investigation of the similarity of Actiwatch counts and Geneactiv-derived counts indicated that the Geneactiv counts were seldom zero for the intervals where corresponding Actiwatch counts were zero. This indicated that a higher sensitivity or higher noise floor was present in the Geneactiv recordings. An optimized threshold thus had to be introduced to correct for this noise floor, which was accomplished as follows. First, Actiwatch epochs were converted to a binary score. All epochs containing any activity (counts  $\geq 1$ ) were scored as 1 and all immobility epochs scored as 0. Similarly, for the Geneactiv recordings all epochs with activity above a given threshold were scored as 1, or 0 if below this threshold. The optimal threshold was defined as the threshold at which there was a minimal absolute difference between the binary scores of both devices. The mean optimal threshold across all Geneactiv recordings of all study participants was subtracted from each epoch and negative counts/epoch were set to zero.

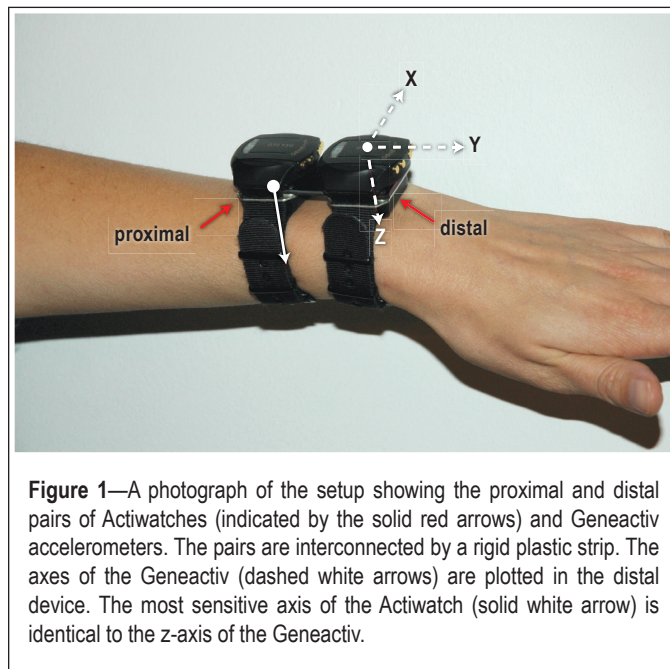
### Optimization of the Transfer Function

Figure 2A shows that preprocessing steps identical to the manufacturer's specification of those taken in the Actiwatch firmware still resulted in substantially smaller Geneactiv-estimated counts per epoch as compared to the Actiwatch counts. To correct for this proportional difference, nonparametric Passing-Bablok regression was used to obtain the optimal regression equation to convert the Geneactiv counts/epoch into Actiwatch counts/epoch.<sup>24,25</sup> The Passing-Bablok method has fewer assumptions regarding the distribution of the data than ordinary least-squares regression. It allows for imprecision in both the reference method (Actiwatch) and the comparison method (Geneactiv), and it is not biased strongly by outliers. For each participant, regression coefficients were calculated between the counts of both distal devices (i.e., GeneactivDist and ActiwatchDist) and both proximal

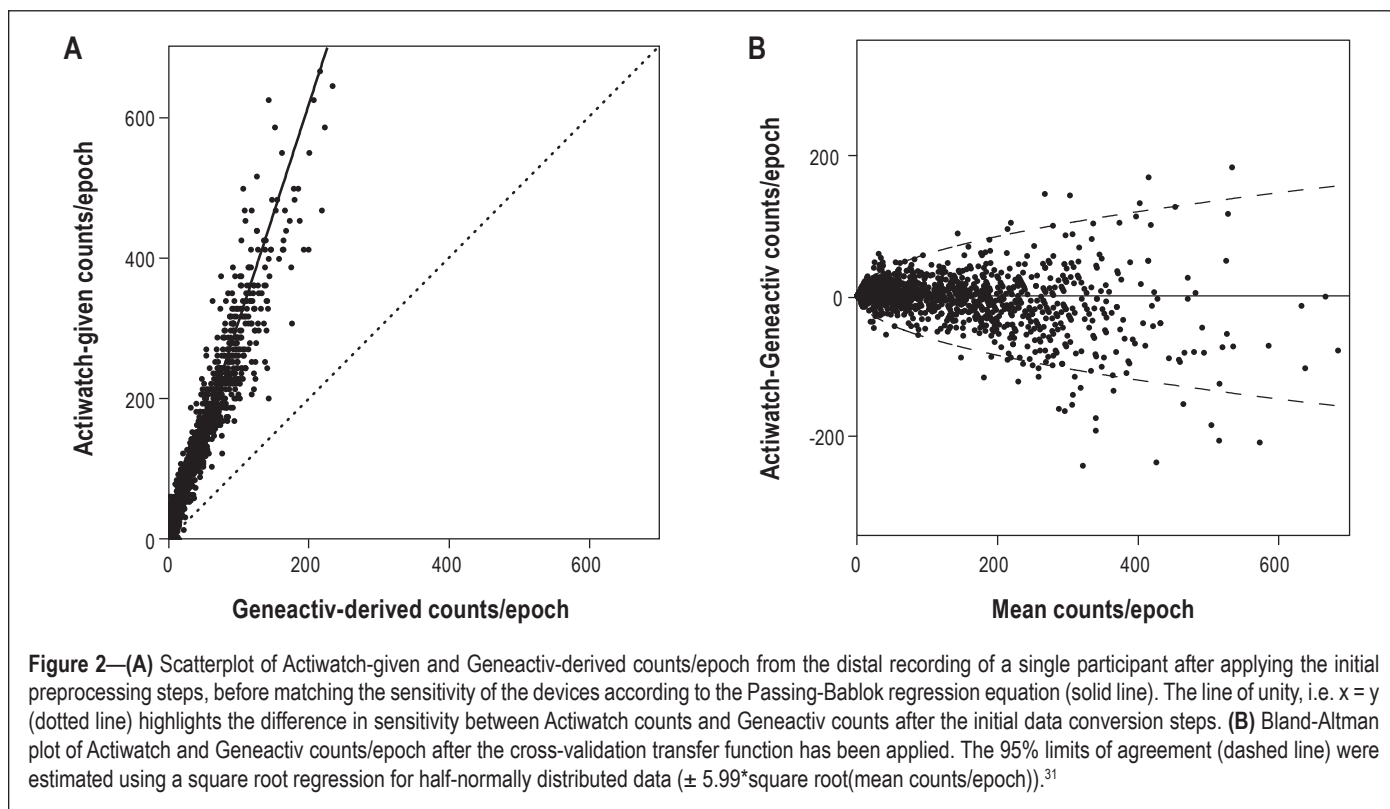
devices (i.e., GeneactivProx and ActiwatchProx), resulting in  $2 \times 15$  ( $2 \times N$ ) regression equations (ActiwatchCount = slope  $\times$  GeneactivCount + intercept). All coefficients were estimated over an equal number of 15-sec epochs (3,600), i.e., the entire recording period from 18:00 until 09:00 on the next day.

### Cross-validation

Preliminary analyses indicated that the regression slopes for signals recorded at the proximal part of the wrist were significantly higher than regression slopes for signals recorded at the distal part of the wrist. Therefore, the  $N$  Passing-Bablok equa-



**Figure 1**—A photograph of the setup showing the proximal and distal pairs of Actiwatches (indicated by the solid red arrows) and Geneactiv accelerometers. The pairs are interconnected by a rigid plastic strip. The axes of the Geneactiv (dashed white arrows) are plotted in the distal device. The most sensitive axis of the Actiwatch (solid white arrow) is identical to the z-axis of the Geneactiv.



**Figure 2**—(A) Scatterplot of Actiwatch-given and Geneactiv-derived counts/epoch from the distal recording of a single participant after applying the initial preprocessing steps, before matching the sensitivity of the devices according to the Passing-Bablok regression equation (solid line). The line of unity, i.e.  $x = y$  (dotted line) highlights the difference in sensitivity between Actiwatch counts and Geneactiv counts after the initial data conversion steps. (B) Bland-Altman plot of Actiwatch and Geneactiv counts/epoch after the cross-validation transfer function has been applied. The 95% limits of agreement (dashed line) were estimated using a square root regression for half-normally distributed data ( $\pm 5.99 \times \text{square root}(\text{mean counts/epoch})$ ).<sup>31</sup>



**Table 1**—Definition of sleep parameters

Time in bed (TIB)	Time between bedtime and get-up time.
Sleep start	The time associated with the first epoch of a 10-min period of immobility starting from lights out time. Within the immobility period 1 epoch of activity was allowed.
Assumed sleep period (ASP)	Time between sleep start and final wake time.
Sleep onset latency (SOL)	Time between lights out time and sleep start.
Total sleep time (TST)	The number of epochs within the assumed sleep period scored as sleep multiplied by the epoch length.
Wake after sleep onset (WASO)	The number of epochs within the assumed sleep period scored as wake multiplied by the epoch length.
Sleep efficiency (SE)	The ratio of total sleep time to assumed sleep time multiplied by 100.
Number of wake bouts (#WB)	The number of continuous blocks of length $\geq 1$ epoch(s) in which each epoch is scored as wake, in the assumed sleep period.
Mean wake bout time (MWBT)	Wake after sleep onset divided by the number of wake bouts.
Number of sleep bouts (#SB)	The number of continuous blocks of length $\geq 1$ epoch(s) in which each epoch is scored as sleep, in the assumed sleep period.
Mean sleep bout time (MSBT)	Total sleep time divided by the number of sleep bouts.

tions were independently cross-validated for the proximal and distal locations using the K-fold approach ('leave-one-out').<sup>26</sup> The mean slope and intercept were calculated over the slopes and intercepts of the equations of *N-1* Geneactiv-Actiwatch pairs of recordings (training pairs). The equation was subsequently applied to convert Geneactiv to Actiwatch counts for the excluded pair of Geneactiv-Actiwatch recordings (validation set). The resulting counts were rounded to the nearest integer. Subsequently, actigraphic sleep parameters were calculated from this series of counts and compared to those obtained from the corresponding Actiwatch data. This was repeated *N* times, such that data from each device pair were used once as a validation set.

### Evaluation of Performance of Derived Counts to Estimate Sleep Parameters

To test the practical applicability of the conversion steps, the algorithm used by the Actiwatch software was applied to the derived Geneactiv counts to estimate sleep parameters (Table 1). The Actiwatch software uses a validated algorithm to classify an epoch as either sleep or wake<sup>6,7</sup>: first, the activity in each epoch is rescored by weighting of activity in the surrounding 2-min period. For each 15-sec epoch the rescored activity is calculated as follows:

$$A_0 = 0.04E_{-(8-5)} + 0.2E_{-(4-1)} + 4E_0 + 0.2E_{+(1-4)} + 0.04E_{+(5-8)}$$

Where  $A_0$  is the total rescored activity for the 15-sec epoch of interest;  $E_0$  is the activity in the scored epoch;  $E_n$  is the activity in the epochs 2 min before ( $E_{-8}$  to  $E_{-1}$ ) and after ( $E_{+1}$  to  $E_{+8}$ ) the scored epoch. If  $A$  is less or equal to a predefined threshold ( $A \leq T$ ) the epoch is scored as sleep, otherwise the epoch is scored as wake ( $A > T$ ). Estimates of the most commonly used sleep parameters, listed in Table 1, were calculated using the low ( $T = 20$  counts), default medium ( $T = 40$  counts) and high threshold setting ( $T = 80$  counts).

### Statistical Analyses

The agreement between individual epochs being scored as wake or sleep by the algorithm was quantified using the Cohen kappa statistic.<sup>27,28</sup> Agreement between all-night sleep parameter estimates derived from Actiwatch counts versus

Geneactiv-derived count estimates was visually inspected with Bland-Altman (mean-difference) plots.<sup>29,30</sup> For each Bland-Altman plot, the assumption of normality of the differences was statistically evaluated using the Shapiro-Wilk test. Due to the half-normal distribution of counts/epoch, the 95% limits of agreement (LOA) in Figure 2B were calculated using regression on the square root of the mean counts/epoch.<sup>31</sup> Spearman rank correlation analysis between the differences and the mean counts/epoch was used to test for constant or proportional bias throughout the measurement range. Inference of the population mean regression slope and the 95% confidence interval (CI) was obtained by applying the bootstrap method. One thousand new mean regression slopes were obtained by resampling with replacement from the original dataset of 15 regression slopes. The CIs were estimated using the bias corrected and accelerated percentile method.<sup>32</sup>

The *a priori* criterion set to allow for the conclusion of adequate performance, the variance between MEMS- and actigraph-derived sleep estimates should ideally be smaller than or equal to the variance between sleep estimates derived from two different exemplars of the same actigraph type.

Actiwatch and Geneactiv data were uploaded to the computer using the Actiwatch Activity and Sleep Analysis software (version 5.08, Cambridge Neurotechnology Ltd., Cambridge, UK) and the Geneactiv PC software (version 1.0, ActivInsights Ltd., Kimbolton, UK), respectively. Data preprocessing, Passing-Bablok regression, cross-validation, bootstrapping and calculation of the Cohen kappa statistic, and sleep parameters were all performed offline using custom-written MATLAB programs (The Mathworks Inc., Natick, Massachusetts, USA). Statistical analyses were conducted using SPSS 18 (SPSS Inc., Chicago, Illinois, USA).

## RESULTS

### Sleep Log

The sleep logs indicated that bed times ranged from 21:30 to 00:31 and lights out times from 21:45 to 00:31. Final wake time and get up times ranged from 05:30 to 08:45 and 05:45 to 09:00, respectively. The average time in bed was  $481 \pm 56$  min (mean  $\pm$  SD).

## Data Conversion Performance

After the Geneactiv z-axis signal had been band-pass filtered, binned, and converted to 15-sec epochs, the derived counts/epoch were converted to binary scores to correct for offset. The mean optimum threshold at which there was a minimal absolute difference between the binary scores of both proximal devices and distal devices was 18 derived counts/epoch (range = [17-19] and [17-18] respectively). After subtracting this offset from all epochs, a linear relation between Actiwatch and Geneactiv counts/epoch was found (Figure 2A). Despite applying the Actiwatch preprocessing specifications to the Geneactiv raw signals, each Geneactiv-derived count/epoch was substantially smaller than its corresponding Actiwatch-given count/epoch.

Linear Passing-Bablok regression was used to calibrate the Geneactiv counts/epoch to Actiwatch counts/epoch. The mean slope of the Passing-Bablok regression (Actiwatch = slope\*Geneactiv + intercept) of the 15 proximal device pairs ( $3.30 \pm 0.19$ ) was significantly higher than the mean slope of the 15 distal device pairs ( $3.07 \pm 0.30$ ):  $0.23 \pm 0.25$ ; 95% CI (0.09; 0.36);  $t(14) = -3.526$ ,  $P = 0.003$ . The intercept was 0 for all 30 regressions due to the high number of zero counts in the data ( $2,385 \pm 249$  and  $2,477 \pm 229$ , for Actiwatch and Geneactiv respectively), which is in agreement with expectation for sleep data. After the regression conversion was applied, the counts/epoch of the Geneactiv were similar to those of the Actiwatch (Figure 2B). Some nonlinearity was observed for > 250 counts/epoch, which did not affect sleep-wake classification because all epochs with such high activity will be scored as wake anyway.

## Agreement of Scoring an Epoch as Sleep or Wakefulness and of Sleep Parameter Estimates

Bootstrapping the mean and CIs of the slope resulted in a mean of  $3.30$  95% CI (3.20; 3.40) and  $3.07$  95% CI (2.95; 3.26) for the proximal and distal slopes, respectively. The agreement of individual epochs being scored as either sleep or wake was almost perfect at all three thresholds (Table 2).<sup>33</sup> When averaged across study participants and distal and proximal placement, kappa values were (mean  $\pm$  SD (range)):  $0.87 \pm 0.05$  (0.75-0.94),  $0.85 \pm 0.06$  (0.71-0.92) and  $0.83 \pm 0.07$  (0.65-0.93) for the low, medium, and high threshold, respectively.

Statistical testing and visual inspection of the Bland-Altman plots revealed one outlier. This was due to a substantial difference in the estimation of sleep onset latency (SOL) between two Actiwatch devices within the same participant (participant 6). This outlier subsequently also affected normality of the differences in the Bland-Altman plots of total sleep time (TST), sleep efficiency (SE), wake after sleep onset (WASO), mean wake bout time (MWBT), and mean sleep bout time (MSBT). It moreover increased the LOA in favor of the newer assessment device (i.e. between-MEMS variance would be smaller than between-actigraph variance). Inspection of the time series of Actiwatch counts revealed that two extra activity peaks postponed the estimation of sleep start in one Actiwatch by 13 min. Although this stresses the need for reliable estimation of SOL, the error was due to a limitation of the sleep scoring algorithm rather than due to the conversion algorithm *per se*. Therefore, a conservative approach was taken and the outlier was excluded from the Bland-Altman

**Table 2**—Percentage of epochs rated as congruent (diagonal) and incongruent (off-diagonal) by Actiwatch and Geneactiv at the medium threshold

	MEMS scores sleep	MEMS scores wake
Actiwatch scores sleep	$91.6 \pm 3.3\%$	$1.0 \pm 0.4\%$
Actiwatch scores wake	$1.0 \pm 0.6\%$	$6.4 \pm 2.9\%$

Data are presented as mean  $\pm$  standard deviation, averaged across all participants and distal and proximal placements. MEMS, microelectromechanical systems.

plots and not used in the estimation of the mean bias and LOA of the differences.

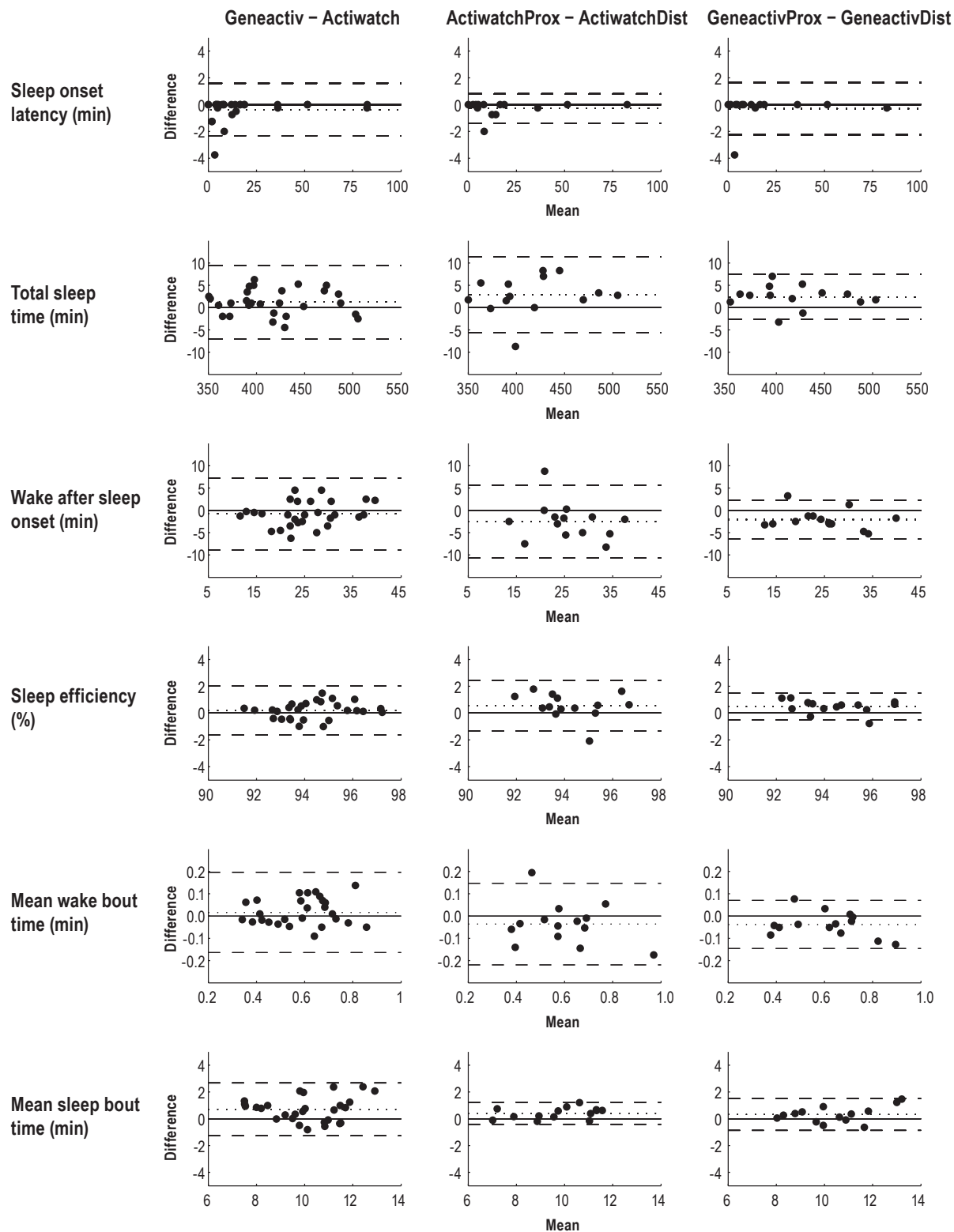
After removal of the outlier, SOL differences were still not normally distributed. However, most of the differences were equal to zero and transformation of the data did not result in normality. Therefore, the estimated LOAs for SOL should be interpreted with caution. The removal of the outlier resulted in normality of the differences of all other sleep parameters and therefore reliable estimates of the LOA (Figure 3).

A nonsignificant Spearman rank correlation between the differences and means of all the Bland-Altman plots confirmed that the bias was uniform throughout the measurement range. At the medium (default) threshold, the bias between two Actiwatch devices was similar to the bias between two Geneactiv devices for SOL, TST, WASO, SE, MWBT, and MSBT (Figure 3).

Only the LOAs of SOL and MSBT were larger between two Geneactiv devices than between two Actiwatch devices, but as stated previously, the LOA of SOL should be interpreted with caution. For all other sleep parameters, the LOAs obtained by two Geneactiv devices were smaller than the LOAs obtained by two Actiwatch devices (Table 3). This indicated more congruent sleep parameter estimates from two Geneactiv devices at two different sites than from two Actiwatch devices at two different sites. The algorithm was also applied using the low and high threshold (data not shown). As sensitivity increased from the high to the low threshold, both the bias and the LOAs increased in magnitude, irrespective of the device. Nevertheless, at all thresholds agreement between two Geneactiv devices was better than between two Actiwatch devices for all sleep estimates, except for SOL and MSBT.

## DISCUSSION

The aim of the study was to investigate whether MEMS-accelerometers could give equivalent sleep parameter estimates as actigraphy. Although this may seem a rather trivial question at first sight, a confirmative answer would have important consequences for sleep research in the current era of large-scale pooled-cohorts GWAS studies. First, because MEMS-accelerometers store the raw and linear acceleration signals, it would become feasible to pool data from different cohorts measured with different brands of MEMS-accelerometers, in contrast to the problematic pooling of data obtained with different brands of actigraphs. Furthermore, a successful conversion algorithm would make it feasible for ongoing long-term cohort or follow-up studies to switch from actigraphs to MEMS-accelerometers



**Figure 3**—Bland-Altman (mean-difference) plots for sleep onset latency, total sleep time, wake after sleep onset, sleep efficiency, mean wake bout time, and mean sleep bout time estimated using the medium threshold of the sleep scoring algorithm. Differences between two measurements by two devices are plotted against the mean of the two measurements. The bias (dotted line) is the mean difference between all measurements by the two devices. The 95% limits of agreement (dashed lines) define the range between which 95% of the differences between measurements by the two devices will lie. The smaller the limits of agreement, the better the agreement between two different devices (between comparison), or between two identical devices (within comparison). The left Bland-Altman plots show the difference in sleep parameter estimation between both proximal and both distal recordings of the two different types of devices (i.e. GeneactivProx-ActiwatchProx and GeneactivDist-ActiwatchDist). These Bland-Altman plots include two recordings per participant, thus the 95% limits of agreement were adjusted for repeated measures.<sup>29</sup> The middle and right Bland-Altman plots show the differences between proximal and distal recording within the same type of device. Please note the difference in the scaling of the axes.

**Table 3**—Summary of bias and 95% limits of agreement of the Bland-Altman plots in Figure 3

	<b>Geneactiv-Actiwatch</b>	<b>ActiwatchProx-ActiwatchDist</b>	<b>GeneactivProx-GeneactivDist</b>
Sleep onset latency (min)	-0.37 (-2.32; 1.59)	-0.29 (-1.39; -0.82)	-0.30 (-2.26; 1.65)
Total sleep time (min)	1.19 (-7.05; 9.43)	2.77 (-5.74; 11.28)	2.39 (-2.62; 7.40)
Wake after sleep onset (min)	-0.82 (-8.85; 7.21)	-2.48 (-10.62; 5.66)	-2.09 (-6.39; 2.22)
Sleep efficiency (%)	0.20 (-1.64; 2.03)	0.55 (-1.33; 2.44)	0.48 (-0.52; 1.48)
Mean wake bout time (min)	0.02 (-0.16; 0.19)	-0.04 (-0.22; 0.15)	-0.04 (-0.15; 0.07)
Mean sleep bout time (min)	0.70 (-1.27; 2.67)	0.41 (-0.41; 1.23)	0.31 (-0.87; 1.50)

Bias and 95% limits of agreement of the Bland-Altman plots of sleep onset latency, total sleep time, wake after sleep onset, sleep efficiency, mean wake bout time, and mean sleep bout time estimated using the medium threshold. The bias is the mean difference between measurements by the two devices. The 95% limits of agreement define the range between which 95% of the differences between measurements by the two devices will lie. The smaller the limits of agreement, the better the repeatability of the device.

without compromising backward compatibility. Finally, the availability of the raw linear acceleration signal can be of use for those studies where the traditional data-reduction and analysis steps do not yield reliable sleep estimates, e.g., in infants<sup>34</sup>; different algorithms can be applied to the stored raw data as they become available.

To design and optimize an algorithm to make MEMS-accelerometers mimic actigraphs, their raw acceleration signal was converted to movement counts, which were subjected to the algorithm used by the Actiwatch software to obtain sleep estimates. Systematic comparisons were made on simultaneous recordings of sets of two Actiwatches and two MEMS-accelerometers. The results indicate that the congruency of sleep estimates obtained from two different MEMS-accelerometer devices was better than the congruency of sleep estimates obtained from two different exemplars of the same type of actigraph. Therefore, most of the disagreement between the Geneactiv and the Actiwatch could be attributed to poor reliability, i.e., poor congruence between two exemplars of the same type of Actiwatch.<sup>29</sup> The transformation algorithm allows MEMS-accelerometers to be used interchangeably with traditional actigraphs, because the 95% LOA for sleep parameter estimates obtained using a MEMS-accelerometer device versus those obtained using a traditional actigraph was equal or better than the 95% LOA for sleep parameter estimates obtained using two exemplars of the same type of actigraph.

Very few studies have addressed the between-device reliability of traditional actigraphs by simultaneously using two exemplars and comparing the sleep parameter estimates generated by the two. Benson et al.<sup>35</sup> compared sleep parameter estimates generated by the simultaneously worn Mini-Motion Logger and Actiwatch during two nonconsecutive sleep recordings. These devices differ with respect to hardware, data reduction algorithms, and scoring algorithms. Unfortunately, no duplicate recordings were made to evaluate the agreement between two different exemplars of the same device type. The authors reported no significant differences between the two types of devices at low and medium sensitivity setting (i.e., a high and a medium threshold, respectively). However, at high sensitivity (i.e., a low threshold), the devices differed significantly on TST, WASO, and SE. Current results show a similar trend of increasing bias and variance as the threshold was lowered.

An interesting observation was that recordings with devices placed more distally on the wrist yielded lower TST and SE and higher WASO than recordings with devices placed more proximally on the wrist. Considering that only the palmar-dorsal axis was analyzed and the larger radius (43 mm) relative to the elbow and shoulder joint, stronger accelerations are expected in the distal recording. Stronger accelerations result in higher counts, likely more epochs with counts beyond the threshold for sleep and thus increased wake detection (WASO), which subsequently decreases TST and SE. Given that small differences in arm length affect sleep estimates, future algorithms based on more sensitive raw accelerometry might consider arm length as an additional variable.

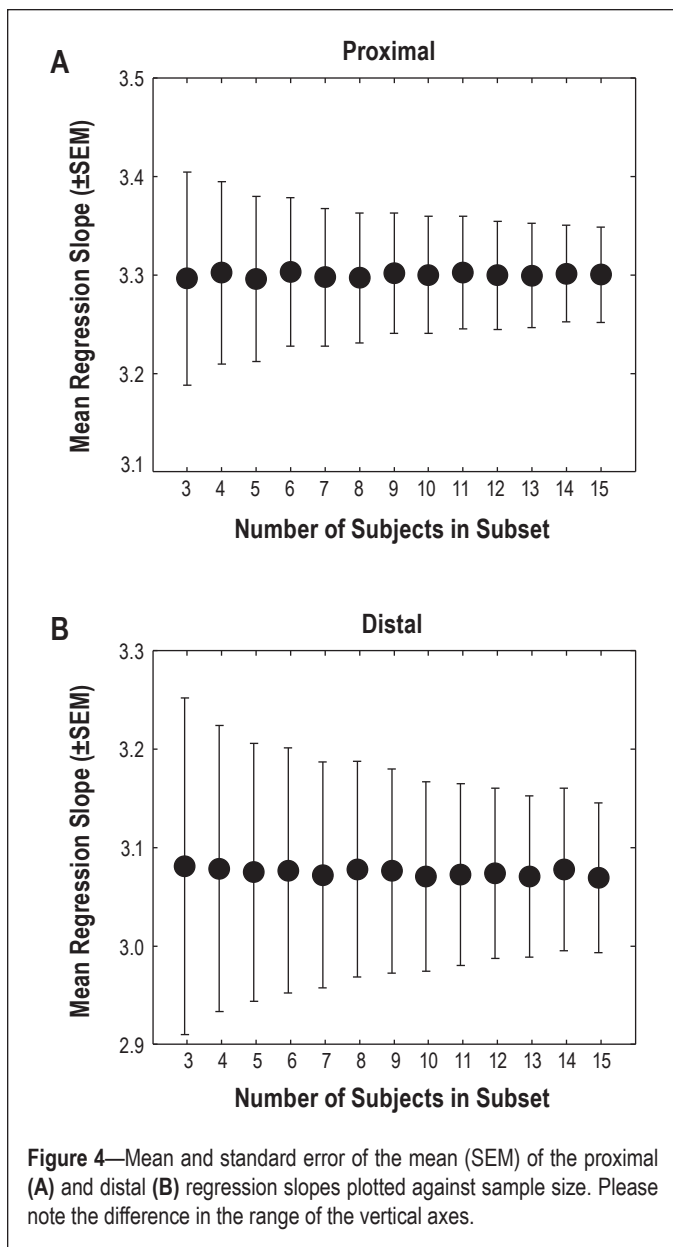
The weight of the Geneactiv devices ( $2 \times 16$  g), the Actiwatches ( $2 \times 17.5$  g), the plastic strip, and the straps was 89 g in total. Thus, the weight of our validation set-up is more than the weight of a single actigraph. However, adding 89 g induces an increase of only 2-3% to the intrinsic weight of the arm and hand.<sup>36,37</sup> At least in our young and healthy population, this finding appears insufficient to considerably affect the accelerations brought about by the upper arm and shoulder muscles. It should be noted that the cumulative weight of the validation setup is comparable to the weight of the first commercially available actigraphs. If cross-validation would be pursued within populations suffering from severe muscle atrophy, the issue would require careful consideration.

Bootstrap analyses on subsamples of our population indicated that follow-up studies can reliably estimate regression coefficients in a sample of 13 (or more) participants (Appendix A).

It should be noted that the conversion algorithm presented here applies to a healthy adult population and is specific for the Actiwatch. Because MEMS record raw accelerometry signals, similar algorithms can be designed for other types of actigraphs that obtain counts by manufacturer-specific preprocessing steps. This results in the further advantage that MEMS recordings can be made compatible with different brands and types of actigraphs.

In conclusion, the current study optimized and validated a data-processing algorithm that makes it possible to pool and exchange data obtained with one of the most frequently used actigraphs and data obtained with MEMS-accelerometers. If anything, more reliable sleep parameter estimates could be obtained from MEMS-accelerometers than from actigraphs,





because MEMS-accelerometers show less variability between different exemplars of the same type of device. The demonstration of validity has strong implications for sleep research. Because MEMS-accelerometers can be produced currently at a fraction of the cost of actigraphs, large-scale assessment of cohorts in GWAS studies, for example, may become feasible. The algorithm provided even makes it feasible for ongoing cohort studies to switch from actigraphy to MEMS-accelerometry while maintaining backward compatibility. To aid in the transition from older Actiwatches to newer MEMS-accelerometry devices, the conversion steps presented in this study were integrated in a MATLAB program with a graphical user interface. The program can be obtained from the authors and allows the user to analyze the data in a similar manner as the Actiwatch Activity and Sleep Analysis software.

The availability of the original raw acceleration signals will not only make it easy to pool data obtained with different brands of MEMS-accelerometers, but also to explore new ways to optimize movement-based sleep estimates. Future

studies that combine video monitoring, PSG, and MEMS-accelerometry can be used to further validate MEMS-accelerometry for sleep-wake discrimination. The possibility for in depth analysis of the three-dimensional acceleration signal might improve the sensitivity and specificity of movement-based sleep estimates. It would be interesting to investigate whether the three-dimensional raw acceleration signal could be exploited to design and validate algorithms to discriminate movements that are specific to sleep stages. For example, not only eye movements<sup>38</sup> but also limb movements<sup>39</sup> during rapid eye movement sleep have been noted to be of a different quality than wake movements. In addition, the availability of the gravitational acceleration may be of use for the detection of the prolonged maintenance of a supine position, which would make the cumbersome and often unreliable logs of bedtime and get-up time superfluous. It may furthermore be used to discriminate gross body movements and postural changes from mere wrist movements and thus improve the estimate of sleep quality.<sup>40</sup> Further improvements of the sleep estimates might be feasible by combining the acceleration signal with other physiological signals, for example skin temperature.<sup>41-43</sup> The availability of MEMS-accelerometers may herald a new era of valid, reliable, and feasible large-scale, field-based assessment of objective sleep estimates.

#### APPENDIX A

For future cross-validation and optimization studies, the number of study participants required to obtain a sufficiently precise estimate of the conversion regression slope should be considered. To provide an indication based on our data, we applied bootstrapping to subsets of the original dataset of 15 regression slopes. For each sample of size  $S$  (between three and 15 study participants) we created 1,000 new datasets of  $S$  regression slopes by sampling, with replacement, from the original dataset of 15 regression slope coefficients. The intercepts of the regressions were all zero and therefore excluded from the analysis. For each of these 1,000 datasets the mean slope was calculated. Subsequently, the mean and standard error of the mean (SEM) of all 1,000 means was calculated and plotted against the sample size (Figure 4). Beyond 13 study participants the reduction in standard error of the mean with every additional study participant appears minimal, indicating that 13 (or more) participants should be sufficient to get a reliable estimate of the mean regression slope.

#### ACKNOWLEDGMENTS

This work was supported by Project NeuroSIPE 10738, of the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs, Agriculture and Innovation; and by the VICI Innovation Grant 453-07-001 of the Netherlands Organization of Scientific Research (NWO); The Hague, the Netherlands. Work for this study was performed at the Department of Sleep and Cognition, Netherlands Institute for Neuroscience.

#### DISCLOSURE STATEMENT

This was not an industry supported study. The authors have indicated no financial conflicts of interest.



## REFERENCES

- Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications. *J Ambulatory Monitoring* 1989;2:209-16.
- Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep* 1992;15:461-9.
- El Baz M, Quera-Salva MA, Oakley NR, Lecendreux M, Gajdos P. Evaluation of actiwatch actimeter vs polysomnography in 29 patients with obstructive sleep apnea syndrome. *J Sleep Res* 1998;7 S2:75.
- Pollak CP, Tryon WW, Nagaraja H, Dzwonczyk R. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep* 2001;24:957-65.
- Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev* 2011;15:259-67.
- Kushida CA, Chang A, Gakary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med* 2001;2:389-96.
- Oakley NR. Validation with polysomnography of the Sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitor system: Technical Report to Mini-Mitter Co., Inc., 1997.
- Chambers MJ. Actigraphy and insomnia: a closer look. Part 1. *Sleep* 1994;17:405-8.
- Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. *Sleep* 1992;15:293-301.
- Hauri PJ, Wisbey J. Actigraphy and insomnia: a closer look, part 2. *Sleep* 1994;17:408-10.
- Van Someren EJW, Van Gool WA, Vonk BFM, Mirmiran M, Speelman JD, Bosch DA, Swaab DF. Ambulatory monitoring of tremor and other movements before and after thalamotomy: a new quantitative technique. *J Neurol Sci* 1993;117:16-23.
- Van Someren EJW, Vonk BFM, Thijssen W, Speelman JD, Schuurman PR, Mirmiran M, Swaab DF. A new actigraph for long-term registration of the duration and intensity of tremor and movement. *IEEE Trans Biomed Eng* 1998;45:386-95.
- Van Someren EJW, Pticek MD, Speelman JD, Schuurman PR, Esselink R, Swaab DF. A new actigraph for long-term tremor recording. *Mov Disord* 2006;21:1136-43.
- Carvalho-Bos SS, Riemersma-van der Lek RF, Waterhouse J, Reilly T, Van Someren EJ. Strong association of the rest-activity rhythm with well-being in demented elderly women. *Am J Geriatr Psychiatry* 2007;15:92-100.
- Gasio PF, Kräuchi K, Cajochen C, Van Someren EJW, Amrhein I, Pache M, Savaskan E, Wirz-Justice A. Dawn-dusk simulation light therapy of disturbed circadian rest-activity cycles in demented elderly. *Exp Gerontol* 2003;38:207-16.
- Van Someren EJ, Scherder EJ, Swaab DF. Transcutaneous electrical nerve stimulation (TENS) improves circadian rhythm disturbances in Alzheimer disease. *Alzheimer Dis Assoc Disord* 1998;12:114-8.
- Vallieres A, Ivers H, Bastien CH, Beaulieu-Bonneau S, Morin CM. Variability and predictability in sleep patterns of chronic insomniacs. *J Sleep Res* 2005;14:447-53.
- Van Someren EJW. Improving actigraphic sleep estimates in insomnia and dementia: how many nights? *J Sleep Res* 2007;16:269-75.
- Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, Thorgerisson TE, et al. A genetic risk factor for periodic limb movements in sleep. *N Engl J Med* 2007;357:639-47.
- Berger AM, Wielgus KK, Young-McCaughan S, Fischer P, Farr L, Lee KA. Methodological challenges when using actigraphy in research. *J Pain Symptom Manage* 2008;36:191-9.
- Matsumoto M, Miyagishi T, Sack RL, Hughes RJ, Blood ML, Lewy AJ. Evaluation of the Actillum wrist actigraphy monitor in the detection of sleeping and waking. *Psychiatry Clin Neurosci* 1998;52:160-1.
- Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENE Accelerometer. *Med Sci Sports Exerc* 2011;43:1085-93.
- Van Someren EJ, Lazeron RH, Vonk BF, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. *J Neurosci Methods* 1996;65:55-62.
- Bablok W, Passing H, Bender R, Schneider B. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *J Clin Chem Clin Biochem* 1988;26:783-90.
- Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 1983;21:709-20.
- Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, 2011 [cited December 15, 2011]. Available from: URL: [http://www.stanford.edu/~hastie/local.ftp/Springer/ESLII\\_print5.pdf](http://www.stanford.edu/~hastie/local.ftp/Springer/ESLII_print5.pdf).
- Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 1960;20:37-46.
- Gale J, Signal TL, Gander PH. Statistical artifact in the validation of actigraphy. *Sleep* 2005;28:1017-8.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135-60.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.
- Bland JM. The half-normal distribution method for measurement error: two case studies, 2005 [cited December 30 2011]. Available from: URL: <http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf>.
- Efron B. Better Bootstrap Confidence-Intervals. *J Am Stat Assoc* 1987;82:171-85.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Insana SP, Gozal D, Montgomery-Downs HE. Invalidity of one actigraphy brand for identifying sleep and wake among infants. *Sleep Med* 2010;11:191-6.
- Benson K, Friedman L, Noda A, Wicks D, Wakabayashi E, Yesavage J. The measurement of sleep by actigraphy: direct comparison of 2 commercially available actigraphs in a nonclinical population. *Sleep* 2004;27:986-9.
- de Leva P. Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. *J Biomechanics* 1996;29:1223-1230.
- Dumas R, Cheze L, Verriest JP. Adjustments to McConville et al. and Young et al. body segment inertial parameters. *J Biomechanics* 2007;40:543-53.
- Aserinsky E. Proportional jerk: a new measure of motion as applied to eye movements in sleep and waking. *Psychophysiology* 1986;23:340-347.
- Chase MH, Morales FR. The atonia and myoclonia of active (REM) sleep. *Annu Rev Psychol* 1990;41:557-84.
- Shimohira M, Shiiki T, Sugimoto J, Ohsawa Y, Fukumizu M, Hasegawa T, Iwakawa Y, Nomura Y, Segawa M. Video analysis of gross body movements during sleep. *Psychiatry Clin Neurosci* 1998;52:176-7.
- Van Someren EJW. Sleep propensity is modulated by circadian and behavior-induced changes in cutaneous temperature. *J Therm Biol* 2004;29:437-44.
- Van Someren EJW. Mechanisms and functions of coupling between sleep and temperature rhythms. *Prog Brain Res* 2006;153:309-324.
- Romeijn N, Van Someren EJW. Correlated fluctuations of daytime skin temperature and vigilance. *J Biol Rhythms* 2011;26:68-77.