International Journal of EXERCISE SCIENCE

Systematic Review

Estimating Physical Activity and Sleep using the Combination of Movement and Heart Rate: A Systematic Review and Meta-Analysis

JAMES W. WHITE III†1, CHRISTOPHER D. PFLEDDERER‡1, PARKER KINARD †1, MICHAEL W. BEETS‡1, LAUREN VON KLINGGRAEFF†1, BRIDGET ARMSTRONG‡1, ELIZABETH L. ADAMS‡1, GREGORY J. WELK‡2, SARAH BURKART‡1, and R. GLENN WEAVER‡1

[1]Arnold School of Public Health, Department of Exercise Science, University of South Carolina, Columbia, SC, USA; [2]Department of Kinesiology, College of Human Sciences, Iowa State University, Ames, Iowa, USA

†Denotes graduate student author, ‡Denotes professional author

ABSTRACT

International Journal of Exercise Science 16(7): 1514-1539, 2023. The purpose of this meta-analysis was to quantify the difference in physical activity and sleep estimates assessed via 1) movement, 2) heart rate (HR), or 3) the combination of movement and HR (MOVE+HR) compared to criterion indicators of the outcomes. Searches in four electronic databases were executed September 21-24 of 2021. Weighted mean was calculated from standardized group-level estimates of mean percent error (MPE) and mean absolute percent error (MAPE) of the proxy signal compared to the criterion measurement method for physical activity, HR, or sleep. Standardized mean difference (SMD) effect sizes between the proxy and criterion estimates were calculated for each study across all outcomes, and meta-regression analyses were conducted. Two-One-Sided-Tests method were conducted to meta-analytically evaluate the equivalence of the proxy and criterion. Thirty-nine studies (physical activity k = 29 and sleep k = 10) were identified for data extraction. Sample size weighted means for MPE were -38.0%, 7.8%, -1.4%, and -0.6% for physical activity movement only, HR only, MOVE+HR, and sleep MOVE+HR, respectively. Sample size weighted means for MAPE were 41.4%, 32.6%, 13.3%, and 10.8% for physical activity movement only, HR only, MOVE+HR, and sleep MOVE+HR, respectively. Few estimates were statistically equivalent at a SMD of 0.8. Estimates of physical activity from MOVE+HR were not statistically significantly different from estimates based on movement or HR only. For sleep, included studies based their estimates solely on the combination of MOVE+HR, so it was impossible to determine if the combination produced significantly different estimates than either method alone.

KEY WORDS: Measurement, validation, calibration, research-grade, accelerometry, sensors

INTRODUCTION

Engaging in healthy levels of physical activity and sleep is critical to leading a healthy life (57, 59, 81). Thus, the accurate assessment of physical activity and sleep is essential for understanding the complex and interdependent relationship between physical activity, sleep, and health outcomes. Scientific-grade sensors, such as ActiGraph, Axivity, Actiheart, and Polar monitors, that use accelerometry to capture movement and electrocardiogram (ECG) or photoplethysmography to capture heart rate (HR), have been validated to assess physical activity and sleep. Calibration/validation studies (37, 76) have estimated physical activity from cut-points or pattern recognition based on movement and demonstrate that accelerometry provides reasonable estimates of physical activity and sleep (79). Studies (4, 9, 15, 71, 84-86) also show that HR provides reasonable estimates of physical activity and classifies time spent in different intensities of physical activity accurately. HR is also one of the essential metrics used in polysomnography (i.e., the gold standard of sleep measurement) to determine sleep and sleep stages (65).

Both measures of movement and HR have limitations that hinder their ability to accurately predict physical activity and sleep. For instance, measures of movement are unable to distinguish between activities that require different energy expenditure from similar movements (e.g., stair climbing, walking uphill) (14, 23). Further, regression equations based on accelerometry in children and adults show overestimated or non-equivalent estimates of energy expenditure for sedentary activities (16, 38), demonstrating that accelerometry cannot accurately capture sedentary activities. Accelerometry also underestimates energy expenditure during intense activities (e.g., fast walking, running) (15, 38). For sleep, accelerometry is limited to only detecting motion (i.e., not sleep), or lack of motion (i.e., sleep). This reliance on movement is likely why accelerometry-based devices for assessing sleep have low specificity and have been shown to over or underestimate sleep time (24, 28). Estimates of motion alone are unable to distinguish sleep stages (18, 19, 28, 79). HR has its own set of limitations, which include the influence of ambient temperature and stress on HR (1, 26, 60). Studies have also shown that HR is influenced by caffeine (35, 75). Devices that use photoplethysmography to assess HR may also be influenced by skin tone (67, 70). While both measures of movement and HR have numerous methodological weaknesses, these two measures are widely used to assess physical activity and sleep.

Given that HR is a physiological indicator and movement is a biomechanical assessment, combining the two methods may address weaknesses of either method alone, which may yield more precise estimates (5). Yet, the benefits of using the combination of movement and HR (MOVE+HR) relative to either method alone has not been meta-analytically quantified in previous literature. Therefore, the aim of this systematic review and meta-analysis was to quantify the difference in physical activity and sleep estimates assessed via 1) movement only, 2) HR only, or 3) MOVE+HR compared to criterion indicators of the outcomes.

METHODS

Participants

This meta-analysis was guided by Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (50). Searches in four electronic databases (PubMed, SportDiscus, Web of Science, and ScienceDirect) were executed September 21-24 of 2021. Search terms were divided into two categories: outcome and design. “Sleep”, “physical activity”, “sedentary”, “circadian”, and “nap” were the search terms used for outcome, and “calib*”, “valid*”, “reliability”, and “accuracy” were the search terms used for design. All possible combinations of the search terms were included using Boolean operators and systematically searched in each database. A full list of search terms and an example search strategy are included in Supplementary Table 1. No restrictions were placed on the year of publication or the age of the population. Studies focused on special populations (e.g., people with a disability, high-level athletes, etc.) were excluded. Titles and abstracts were searched using Medical Subject Heading Terms (MeSH-terms) for databases using this indexing and cataloging structure (e.g., PubMed). For databases that did not use MeSH-terms for their indexing and cataloging structure (i.e., SportDiscus, Web of Science, and ScienceDirect), titles and abstracts were searched instead of searching the full text of articles. Citations of articles that were included after the data management and selection process were screened for relevant articles. This research was carried out fully in accordance to the ethical standards of the International Journal of Exercise Science (53).

Inclusion and Exclusion Criteria

Peer-reviewed, published validation/calibration studies of human participants were included if they 1) estimated physical activity outcomes, movement behaviors, or sleep using the combination of movement and HR collected via a research-grade measurement tool, and 2) used a criterion indicator of physical activity, movement behaviors, or HR as a reference for estimates from proxy measurement tools. Research-grade measurement tools were defined as tools that are primarily marketed to research scientists for use in research studies (6, 34). Criterion measures are defined as the most direct indicators of the outcome, in this instance physical activity and sleep. Criterion measurement methods in this study included indirect or direct calorimetry, doubly labeled water, direct observation, electrocardiogram, and polysomnography. Proxy measures are indirect indicators of a desired outcome that are closely related to that outcome. For this study, proxy measurement methods included movement, HR, or MOVE+HR. Only studies printed in English were included in the review.

Animal studies and computer-simulated models were excluded, along with retracted studies, letters to the editor, and conference abstracts. Studies that only used movement or HR to estimate physical activity or sleep were excluded because they would increase the heterogeneity across studies, as opposed to analyzing within-study comparisons of movement or HR only and/or the combination of MOVE+HR. Studies that included movement and HR in addition to several other metrics (e.g., eye movement, muscle activity, blood saturation, breathing rate, etc.) were excluded. Additionally, studies that predicted sleep disorders (e.g., sleep apnea, restless leg syndrome, and leg movement), did not evaluate a device worn by participants (not the case for sleep studies, as most sleep occurs in the bedroom), evaluated special populations, or evaluated consumer wearable devices as measurement tools were also excluded. Consumer wearable devices were considered devices that were primarily marketed to the public as end-users to monitor health metrics (e.g., Fitbit, Garmin).

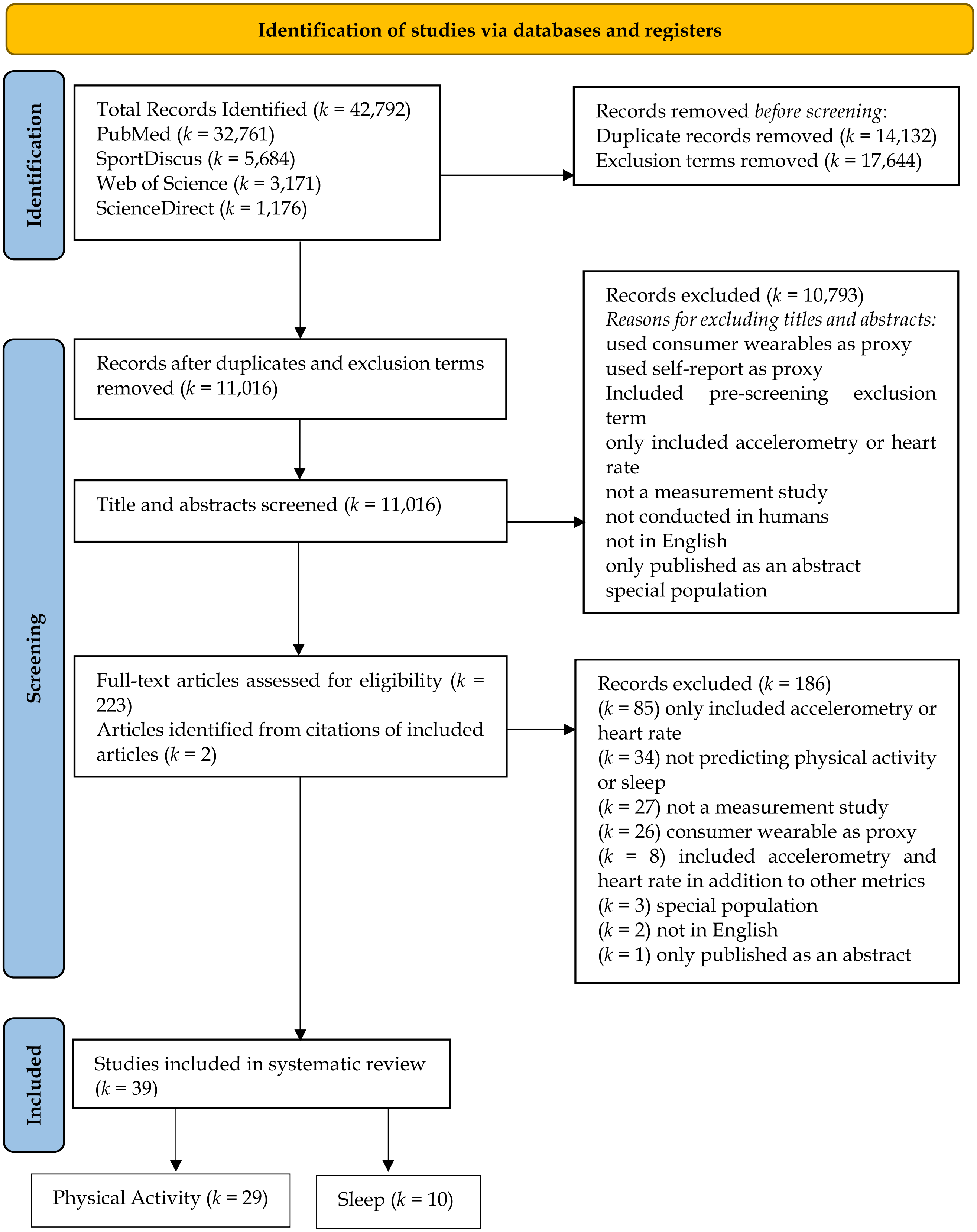


Figure 1. Prisma flow diagram.

Study Records

The studies identified in the searches were screened for inclusion in a three-step process. First, duplicate articles in the resulting searches were removed and studies were pre-screened in Endnote (Clarivate, London, United Kingdom). The EndNote "Find Duplicates" function was used to remove duplicate articles from within each database. Because of the large number of search results, pre-identified exclusion (e.g., survey, questionnaire, etc.) and inclusion (e.g., Actiheart, Polar) terms selected by the study team were applied using the "simple search" function in EndNote to pre-screen irrelevant studies prior to reviewing titles and abstracts. A full list of exclusion and inclusion terms is provided in Supplementary Table S2. Second, the remaining titles and abstracts were screened in Covidence, an online systematic observation management tool. Four independent reviewers screened titles and abstracts for full-text eligibility using the same criteria established for the initial search strategy. Any inconsistencies or questions related to the inclusion of a study were resolved via consensus with an additional author. Third, once title and abstract screening was completed, the full-text articles were retrieved and reviewed for inclusion using the same criteria established for the initial search strategy by two independent reviewers. Again, any inconsistencies or questions related to the inclusion of a study were resolved via consensus with a third author. Consensus was reached on all studies included.

Data Extraction and Coding

Data extraction was conducted by two authors. The extracted data were placed in a custom extraction spreadsheet created in Microsoft Excel. To ensure consistency of extraction, authors independently extracted the first 9 studies to be included for meta-analysis, then met with an author to resolve any inconsistencies or questions related to the extraction of a study. All three authors met weekly to resolve discrepancies via consensus. If authors could not come to a consensus in weekly meetings, additional authors provided further input. Through this process, consensus was reached on all included articles.

General study characteristics were extracted from all studies and comprised of the following variables: study author, publication year, study title, overall sample size, number of female participants, sample characteristics (e.g., age, ethnicity), and country in which the study was conducted. Physical activity or sleep metric predicted (i.e., energy expenditure, physical activity energy expenditure, metabolic equivalents, total sleep time, sleep stage, sleep efficiency, body posture), proxy signal used for prediction (i.e., HR, movement, or combined MOVE+HR), proxy measurement tool (e.g., ActiGraph accelerometer, Polar HR monitor), placement of proxy measurement tool, and criterion measurement method (i.e., indirect calorimetry, direct calorimetry, doubly labeled water, and polysomnography) were also extracted for both physical activity and sleep studies. Effect estimates (e.g., mean, median) and variability of those estimates (e.g., standard deviation, interquartile range, standard error) were extracted. Other presented metrics indicating the validity of the proxy compared to the criterion were also extracted (e.g., % accuracy, r-squared of prediction equation, mean error, mean absolute error).

Assessment of Methodological Quality

Two reviewers independently examined the risk of bias for each included study based on recommendations about selecting appropriate accelerometer cut-points for youth from a previous systematic review (37). Studies were examined for quality using four criteria: 1) use of an appropriate criterion measure of physical activity or sleep, 2) a wide variety of activities were included in the protocol (i.e., at least 6 activities), 3) accelerometer data was collected with epochs of 60 seconds or less, and 4) had more than 10 participants based on the study's predefined age group(s) (e.g., Zakeri 2013 reported greater than 10 participants for 3-, 4-, and 5-year age groups). Criteria 1-3 were rated as "yes", "no", or "could not determine" for each of the included studies (see Supplementary Tables 3 and 4). Criterion 4 was rated using a two-step process. First, we identified study sample size and then second, we rated the study "yes" if the study had more than 10 participants based on the study's predefined age group(s) and "no" if the study had less than or equal to 10 participants based on the study's predefined age group(s).

Table 1. Demographics of the included physical activity studies.

Study	Sample Size	*N* Female (%)	Sample Age Mean (SD) or Sample Age Range[a]	Metric Predicted	Criterion Measure	Proxy Signal	Proxy Measure (Location)
Hernandez-Vicente, 2016(32)	18	9 (50)	21.0 (1.2)	EE	ACC	Combo	Polar (Wrist)
Rennie, 2000(61)	8	3 (37)	23-54	EE	DC	Combo, HR	Custom Device (Chest)
Moon, 1996(51)	20	10 (50)	19-40	Oxygen Consumption, Carbon Dioxide Production	DC	Combo, HR	Polar (Chest, Wrist), Mini-mitter (Leg)
Butte, 2014(10)	50	25 (50)	4.5 (0.8)	EE	DLW	Combo, ACC	Actigraph (Hip), Actiheart (Chest)
Silva, 2015(71)	17	0 (0)	24.9 (4.8)	EE, PAEE	DLW	Combo, ACC, HR	Actiheart (Chest)
Ojiambo, 2012(56)	49	25 (51)	6.9 (1.5)	EE, PAEE	DLW	Combo, ACC	3DNX v3 (Hip), ActiGraph (Hip), Sunto t6 (Chest)
Santos, 2014(66)	12	8 (67)	16.4 (0.5)	EE, PAEE	DLW	Combo, ACC, HR	Actiheart (Chest)
Garnotel, 2018(25)	56	25 (45)	39.6 (12.7)	EE, PAEE	DLW	Combo, ACC	Actigraph (Hip),

							Actiheart (Chest)
Villars, 2012(80)	35	0 (0)	28.4 (8.4)	PAEE	DLW	Combo, ACC	Actiheart (Chest), RT3 (Wrist)
Assah, 2011(2)	33	9 (27)	35.4 (6.8)	PAEE	DLW	Combo, ACC, HR	Actiheart (Chest)
Campbell, 2012(11)	18	10 (55)	17.5 (0.6)	PAEE	DLW	Combo	Actiheart (Chest)
Lof, 2013(45)	20	20 (100)	36.0 (8.0)	PAEE	DLW	Combo, ACC	Actigraph (Hip), Actiheart (Chest), IDEEA (Waist)
Bussman, 1998(8)	3	0 (0)	19-24	Activities detected	DO	Combo	IC-3031 (Thigh), Vitaport recorder (Chest)
Zakeri, 2008(84)	109	44 (40)	12.3 (3.5)	EE	IC	Combo	Actiheart (Chest)
Zakeri, 2010(85)	61	26 (43)	11.8 (3.8)	EE	IC	Combo	Actiheart (Chest)
Zakeri, 2013(86)	69	35 (51)	4.6 (1.0)	EE	IC	Combo, ACC	Actiheart (Chest), Actigraph (Hip)
Rumo, 2011(64)	44	16 (36)	35.0 (11.0)	EE	IC	Combo	Custom Device (Chest, Wrist, Ankles, Back, Arms)
Spierer, 2011(72)	27	11 (41)	26.3 (7.3)	EE	IC	Combo, ACC, HR	Actical (Hip), Actiheart (Chest)
Barreira, 2009(3)	34	17 (50)	21.8 (3.6)	EE	IC	Combo, ACC	Actigraph (Hip), Actiheart (Chest)
Ellis, 2014(21)	40	21 (52)	35.8 (12.1)	EE	IC	Combo, ACC	Actigraph (Wrist, Hip), Polar (Wrist)
Strath, 2001(73)	30	14 (47)	32.5 (12.7)	METs	IC	Combo, ACC, HR	Polar (Wrist), CSA (Hip, Wrist, Thigh), Yamax (Hip)

Schneller, 2015(68)	14	8 (57)	27.7 (3.3)	METs	IC	Combo, ACC	Actigraph (Hip), Actiheart (Chest), ActivPal (Thigh)
O'Driscoll, 2021(55)	89	54 (61)	38.2 (14.1)	METs	IC	Combo, ACC	Actigraph (Hip), Polar H7 (Chest), Sensewear (Arm)
Eston, 1998(22)	30	15 (50)	9.3 (0.8)	Oxygen Consumption	IC	Combo, ACC	WAM, BHL 6000 Medical (Chest), Tritrac-R3D (Hip), Digimax (Hip)
Corder, 2005(14)	39	16 (41)	13.2 (0.3)	PAEE	IC	Combo, ACC, HR	Actigraph (Hip, Ankle), Actical (Hip), Actiheart (Chest)
Brage, 2004(5)	12	0 (0)	22.7-30.6	PAEE	IC	Combo, ACC, HR	Actigraph (Hip), Polar (Wrist)
Brage, 2005(4)	20	9 (45)	26-50	PAEE	IC	Combo, ACC, HR	Actiheart
Corder, 2007(13)	145	79 (54)	12.4 (0.2)	PAEE	IC	Combo, ACC	Actigraph (Hip), Actiheart (Chest)
Gilgen-Ammann, 2021(27)	23	0 (0)	20.9 (2.5)	PAEE	IC	Combo	Actiheart (Chest), Equivital Hidalgo EQ02 (Chest), Everion (Arm), PADIS 2.0 (Chest)

Abbreviations: "EE" Energy Expenditure, "PAEE" Physical Activity Energy Expenditure, "METs" Metabolic Equivalents of Task, "ECG" Electrocardiogram, "IC" Indirect Calorimetry, "DC" Direct Calorimetry, "DLW" Doubly Labeled Water, "DO" Direct Observation, "HR" Heart Rate, "ACC" Accelerometry, "Combo" Combination of Heart rate and Accelerometry

[a]When mean and standard deviation are not reported

Table 2. Demographics of the included sleep studies.

Study	Sample Size	*N* Female (%)	Sample Age Mean (SD) or Sample Age Range[a]	Metric Predicted	Criterion Measure	Proxy Signal	Proxy Measure (Location)
Takano, 2018(74)	7	0 (0)	20-23	Body Posture	Biopac Bionomadix Wireless RSP with ECG	Combo	FUSE bed sheet sensor (Under Bed)
Yoshihi, 2021(83)	8	0 (0)	21.8 (0.7)	Sleep Stages	PSG	Combo	Polymate Pro KXM52-105 accel (Head)
Kortelainen, 2010(39)	9	9 (100)	20-54	Sleep Stages	PSG	Combo	Emfit bed sensor (Under Bed)
Herscovici, 2006(33)	30	9 (30)	46.8 (14.8)	Sleep Stages	PSG	Combo	Itamar WP_100 PAT recorder (Wrist)
Kuula, 2021(40)	20	10 (50)	24.5 (3.1)	Sleep Stages	PSG	Combo	Firstbeat Bodyguard 2 (Chest), Geneactive (Wrist)
Mitsukura, 2020(49)	25	11 (44)	NR	Sleep Stages	PSG	Combo	BCG Sensors (Under Bed)
Mendez, 2009(47)	6	6 (100)	40-50	Sleep Stages, Sleep Efficiency	PSG	Combo	Emfit bed sensor (Under Bed)
Migliorini, 2010(48)	11	11 (100)	NR	Sleep Stages, Sleep Efficiency	PSG	Combo	Emfit bed sensor (Under Bed)
Muzet, 2016(52)	12	6 (50)	18-40	Sleep Stages, Sleep Efficiency, Total Sleep Time	PSG	Combo	Actigraph (Wrist), ECG (Chest)
Da Woon, 2014(17)	20	4 (20)	38.7 (14.6)	Sleep Stages, Sleep Efficiency, Total Sleep Time	PSG	Combo	PVDF Film Sensor -BCG (Above Mattress)

Abbreviations: "NR" Not Reported, "PSG" Polysomnography, "ECG" Electrocardiogram, "Combo" Combination of Heart Rate and Accelerometry

[a]When mean and standard deviation are not reported

Statistical Analysis
Descriptive statistics of the included studies were calculated prior to analyses. A two-step approach to estimate the group and individual differences between proxy and criterion estimates was taken due to concerns of ecological fallacy which may be present in measurement studies (4). In other words, an estimate of physical activity and/or sleep produced by a proxy may be comparable to the criterion at the group level, while no estimates of physical activity and/or sleep for any individual in that group are accurate. To give a general example, in one validation study, a proxy measure may be -10 units away from the criterion for one participant, but +10 units away from the criterion for another participant, resulting in an average difference of 0 units from the criterion. To account for this, study/group level percent error (MPE_{group}) and mean absolute percent error ($MAPE_{group}$) of the proxy compared to the criterion measure were calculated first, using the following formula:

$$\left(\frac{\text{criterion estimate} - \text{proxy estimate}}{\text{criterion estimate}}\right) * 100$$

This calculation was applied for studies that presented a group-level point estimate of the proxy and criterion measure. This provided a standardized (i.e., not in the measured units) study/group level estimate of the MPE_{group} and $MAPE_{group}$ of proxy compared to criterion. The weighted mean (by sample size) was then calculated across studies. Second, a weighted mean for studies that presented mean absolute percent error ($MAPE_{ind}$) was calculated across studies separately by proxy signal (i.e., movement only, HR only, and MOVE+HR). The $MAPE_{ind}$ provides a metric of how well the proxy predicts the criterion for each participant in these studies. Weighted means were also calculated for r-squared and percent accuracy because these metrics do not include variability around the point estimates, which are required for pooling study estimates in traditional meta-analytic approaches.

Following the approach above, a more traditional meta-analysis approach was completed. Standardized mean difference (SMD) effect sizes between the proxy and criterion estimates were calculated for each study across all outcomes in Comprehensive Meta-Analysis (v.3.0). Meta-regression analyses were conducted in Stata (v.16.1, StataCorp, College Station, Texas) using the 'robumeta' command that uses robust variance estimation methods in order to account for the clustered estimates within studies (31). All models included the SMD between proxy and criterion as the dependent variable and dummy variables representing the proxy signals as the independent variables. Study sample size, percent female, location of proxy measurement device (i.e., chest, wrist, combination), and criterion measurement method (i.e., direct calorimetry, indirect calorimetry, doubly labeled water, and polysomnography) were included as covariates.

Finally, equivalence tests using the Two-One-Sided-Tests method (69) were conducted to meta-analytically evaluate the equivalence of the proxy and criterion (62). To reject the null hypothesis that proxy and criterion are not equivalent, the 90% confidence interval for the difference between proxy and criterion is required to fall within prespecified equivalence bounds (20). Consistent with best practices in meta-analyses (41, 42), equivalence bounds were set based on a range of SMD sizes from small to large (0.2, 0.4, 0.6, 0.8). Following the primary analyses, secondary subgroup analyses were conducted for children (i.e., < 19 years) and adults (≥ 19 years) and by criterion measurement method (i.e., direct calorimetry, indirect calorimetry, and doubly labeled water) following the same meta-regression and equivalence testing procedures. All models were conducted using random effects weighting schemes (78).

RESULTS

Search Results

Figure 1 presents the PRISMA flow diagram. A total of 42,792 articles were identified from the database searches with 11,016 articles remaining after duplicates were removed and pre-screening exclusion and inclusion terms were applied. A total of 10,793 articles were excluded based on title and abstract, leaving 223 articles for full-text review. Two additional articles were identified from citations of included articles. After full-text screening, 186 studies were excluded, leaving 39 articles for data extraction. A total of 29 articles were physical activity studies and 10 were sleep studies.

Study Characteristics

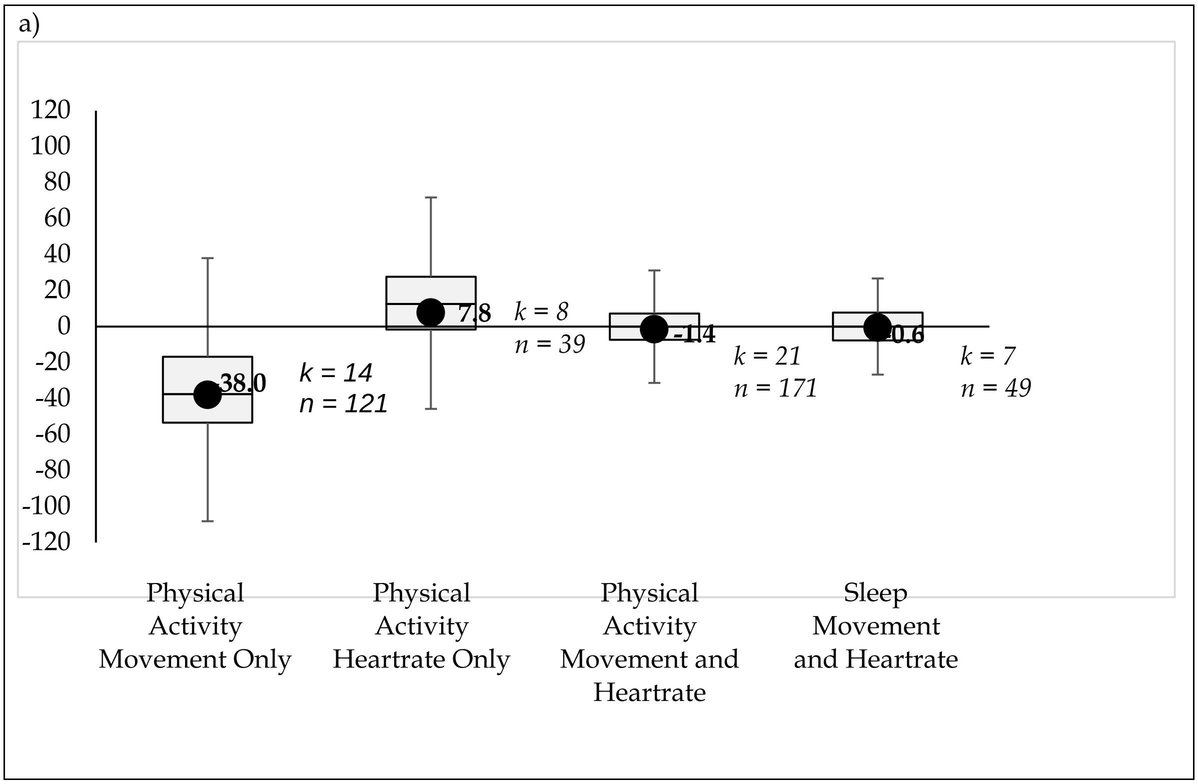
This review included a total of (n = 1,273) participants, with physical activity studies including (n = 1,125) participants and sleep studies including (n = 148) participants. Characteristics of the included physical activity studies are presented in Table 1 and sleep studies in Table 2. Most physical activity studies (k = 20) and sleep studies (k = 7) reported 50% or greater male participants. Race of participants was reported in few studies for physical activity (k = 6) or sleep (k = 0). For physical activity, most studies were conducted in adults (k = 19). All studies for sleep were conducted in adults. The most common physical activity metric predicted was physical activity energy expenditure (k = 13), followed by overall energy expenditure (k = 14), metabolic equivalents (k = 3), oxygen consumption (k = 2) and types of activities (k = 1). For sleep, the most common metric predicted was sleep stage (k = 9), followed by sleep efficiency (k = 4), total sleep time (k = 2), and body posture (k = 1). For physical activity, the most common criterion measurement method was indirect calorimetry (k = 16), followed by doubly labeled water (k = 9), direct calorimetry (k = 2), direct observation (k = 1), and accelerometry (k = 1). For sleep, polysomnography was used as the criterion measurement method in all but one study which used Biopac Bionomadix Wireless Dual Wireless Respiration and Electrocardiogram.

Quality Assessment

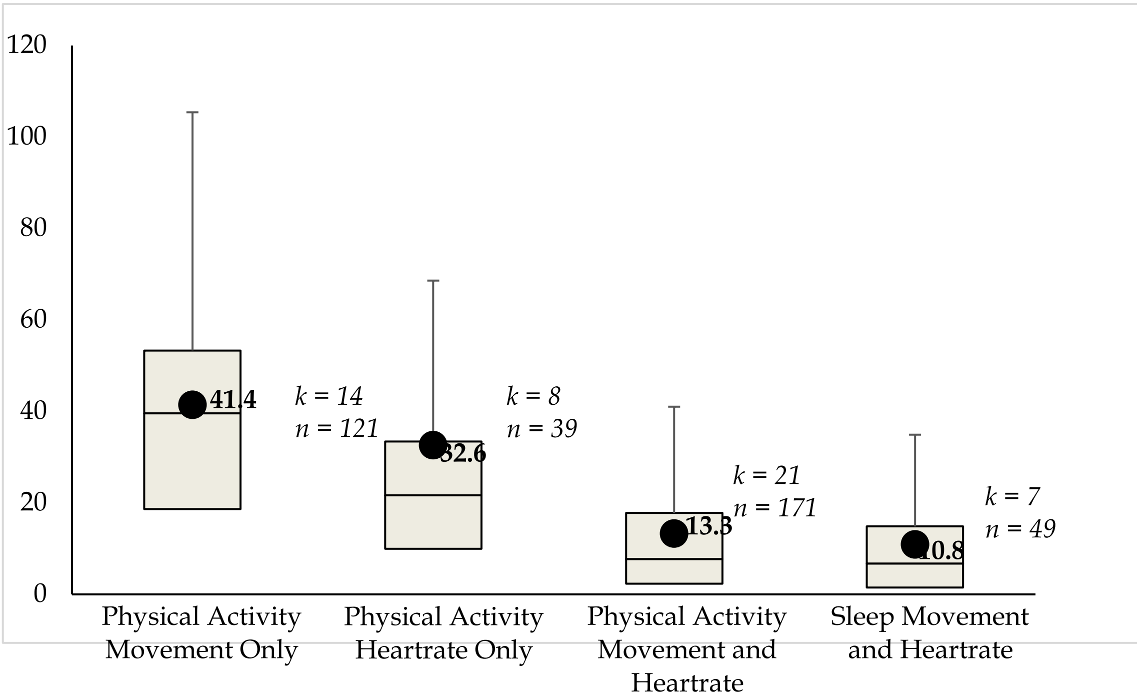
Findings from the quality assessment of included studies are presented in Supplementary Tables 3 and 4. Nearly all studies used an appropriate measure of physical activity or sleep (k = 37), while fewer studies included a wide variety of activities (k = 15), collected data with an epoch of less than 60 seconds (k = 21), and/or had more than 10 participants for each age group (k = 6).

Meta-Analytic Findings

Figure 2 presents box plots for physical activity and sleep indicating the median, interquartile ranges, and the weighted mean of the MPE_{group}, $MAPE_{group}$, $MAPE_{ind}$, r-squared, and percent accuracy by proxy signal and behavior compared to criterion. For physical activity movement only, HR only, and MOVE+HR, weighted means for MPE_{group} were -38.0%, 7.8%, and -1.4%, respectively, compared to the criterion. For physical activity movement only, HR only, and MOVE+HR, weighted means for $MAPE_{group}$ were 41.4%, 32.6%, and 13.3%, respectively, compared to the criterion. Weighted mean for $MAPE_{ind}$ was 26.6% for MOVE+HR compared to the criterion. For physical activity movement only, HR only, and MOVE+HR weighted means for r-squared were 0.52, 0.64, and 0.64, respectively, compared to the criterion. For sleep MOVE+HR, weighted mean for MPE_{group} was -0.6% compared to the criterion. For sleep MOVE+HR, weighted mean for $MAPE_{group}$ was 10.8% compared to the criterion. For sleep MOVE+HR, weighted mean for percent accuracy was 80.5% compared to the criterion.

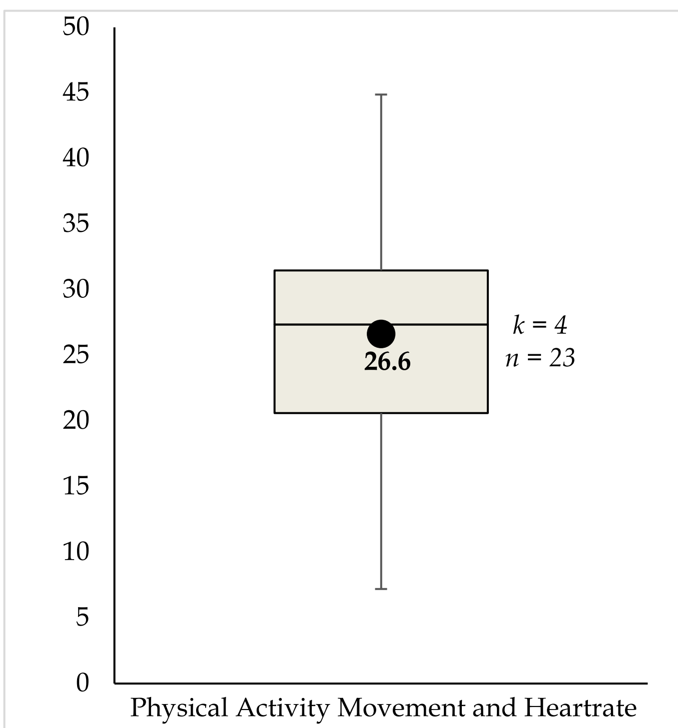


b)



120
100
80
60
40
20
0
41.4
k = 14
n = 121
32.6
k = 8
n = 39
13.3
k = 21
n = 171
10.8
k = 7
n = 49
Physical Activity Movement Only
Physical Activity Heartrate Only
Physical Activity Movement and Heartrate
Sleep Movement and Heartrate

c)



50
45
40
35
30
25
20
15
10
5
0
26.6
k = 4
n = 23
Physical Activity Movement and Heartrate

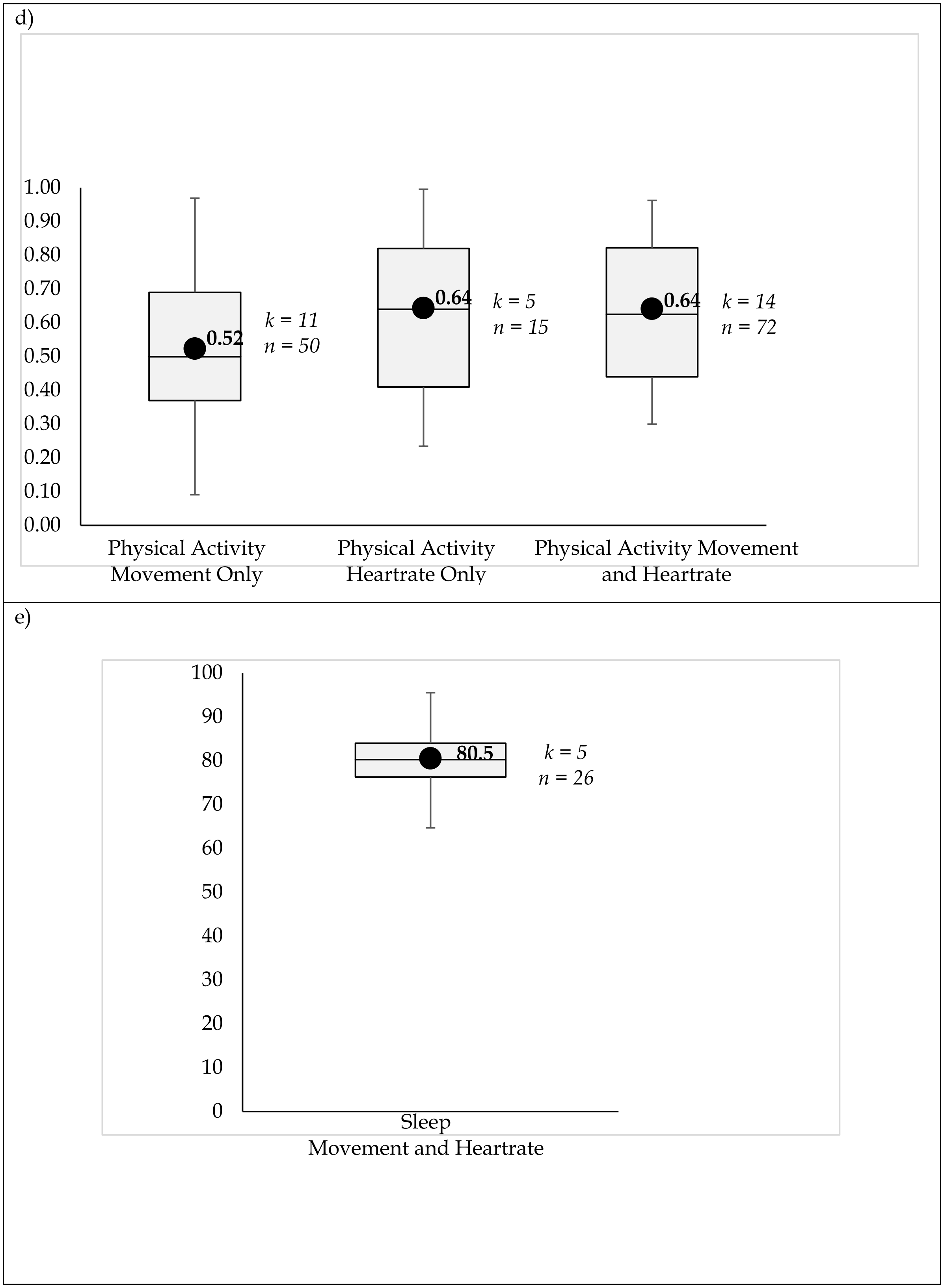
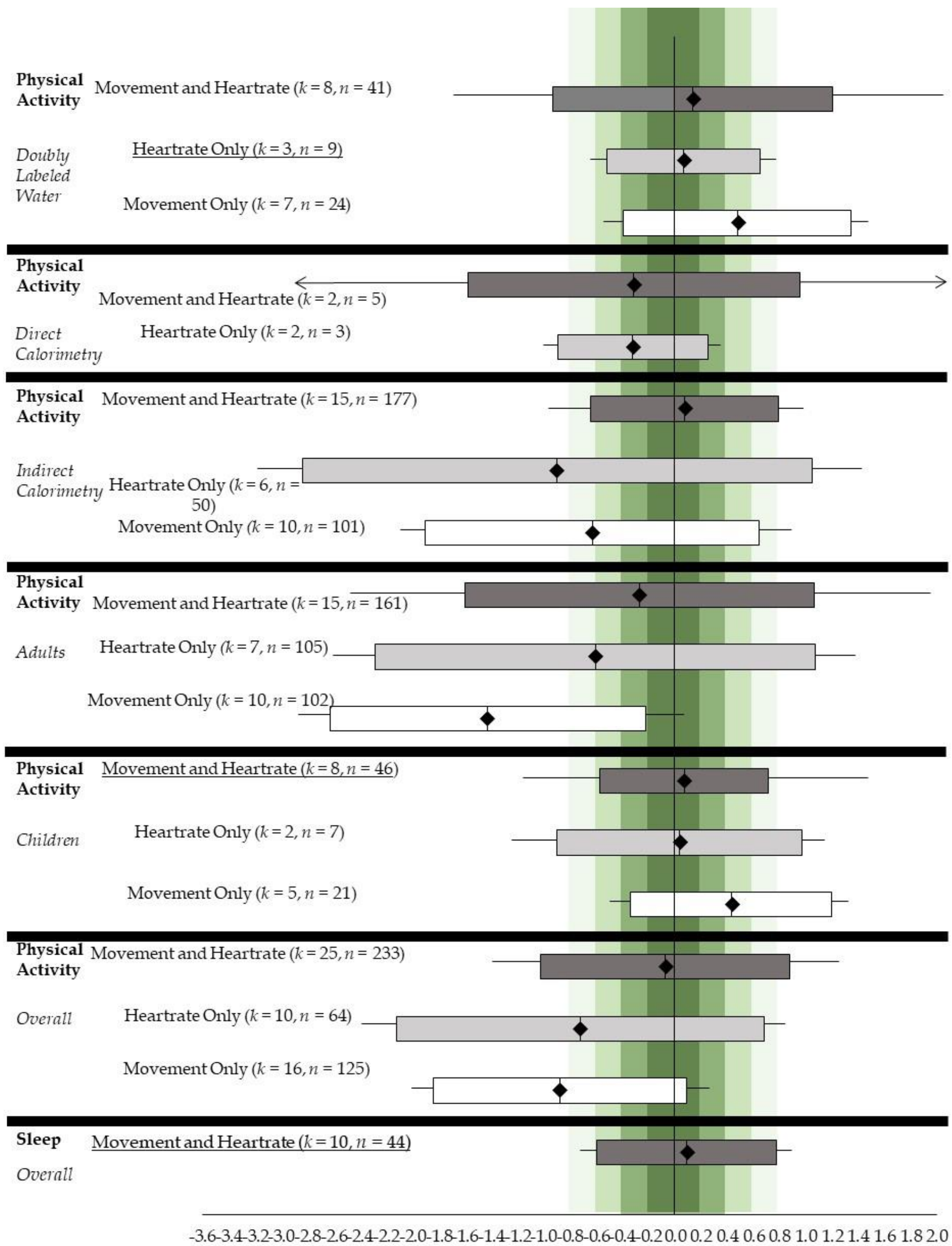


Figure 2. Box Plots of a. percent error calculated from group means, b. absolute percent error calculated from group means, c. study reported mean absolute percent error, d. r-squared, and e. percent accuracy by proxy signal and behavior compared to the criterion.

Summary effect sizes and equivalence bounds from the meta-regressions estimating the SMDs are presented in Figure 3. None of the primary or secondary analyses showed statistically significant differences between the proxy and criterion. However, only three analyses indicated that the proxy and criterion were statistically significantly equivalent at a SMD of 0.8. These analyses included MOVE+HR when estimating sleep, MOVE+HR when estimating physical activity in children, and HR alone when compared to the criterion of doubly labeled water for estimating physical activity.



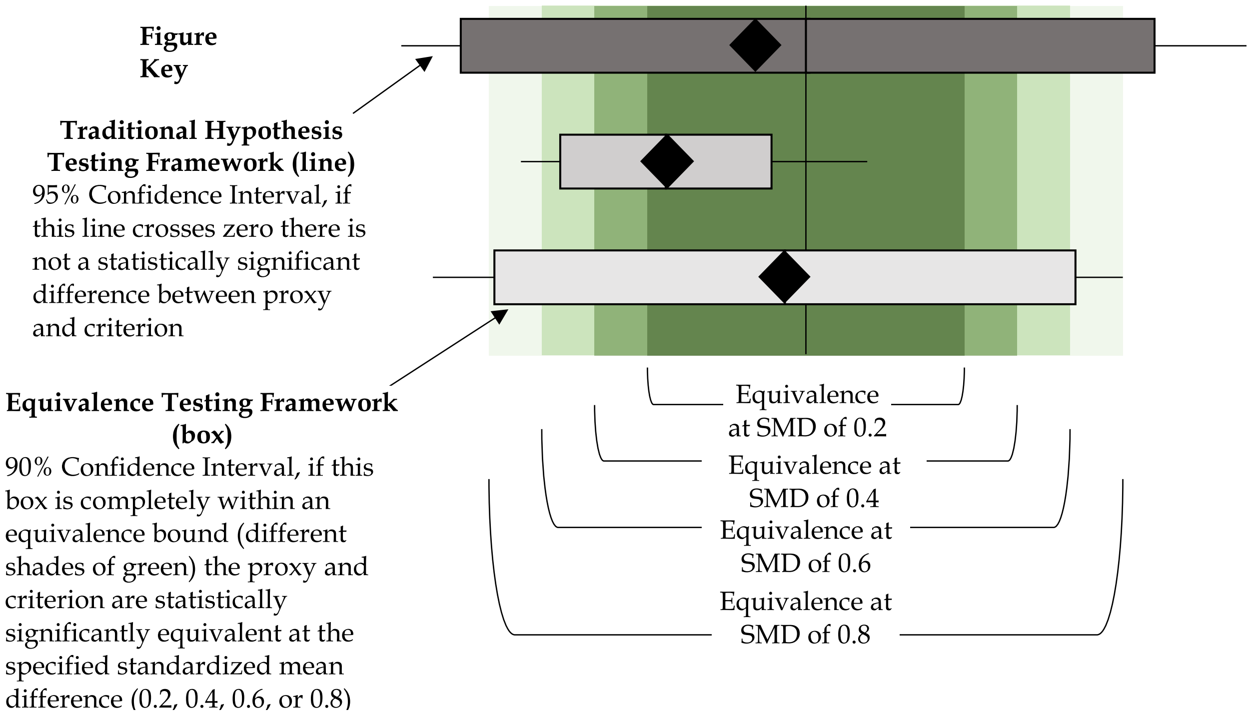


Figure 3. Summary Effect Sizes and Equivalence Bounds. No analyses showed that proxy and criterion were statistically sigificantly different in a traditional hypothesis testing framework at a $p < 0.05$. Underlined analyses were statistically significantly equivalent at a standardized mean diffence of 0.8.

While not statistically significant, a clear pattern emerged where MOVE+HR improved estimates of physical activity compared to movement or HR only. A similar systematic review and meta-analysis that evaluated the accuracy of energy expenditure estimates from wrist- or arm-worn devices during different activity types in adults reported that the accuracy of energy expenditure estimates improved with the addition of heart rate sensing during most activities (54). The lack of statistical difference between estimates of physical activity based on the proxy signals compared to the criterion may have been due to the large heterogeneity across the included studies. Sources of heterogeneity included the wide variety of conditions (i.e., laboratory, semi-structured, and free-living), measurement tools, device placements, etc. (see Table 1). Recent best practice guidelines for evaluating new physical activity and sleep measurement devices have been published (36, 82, 83). Future studies that comply with these guidelines may reduce study heterogeneity in future systematic reviews.

Across the included physical activity studies, another pattern emerged where the use of movement alone to predict physical activity was consistently the least accurate, followed by HR. This pattern was consistent across children and adults and different criterion methods. Notably, this finding is relevant because movement, not HR, is currently the most used method for assessing free-living physical activity (46, 76, 77). This is likely because, until recently, movement was easier to assess compared to HR—HR telemetry or wireless transmission of HR signals via a chest strap was more costly and burdensome (for both researchers and participants) than measurement tools that assessed movement (63). Further, many methods of estimating physical activity via HR require individual calibration, while estimates of physical activity based on movement do not require the extra step of individual calibration (13, 44). However, the emergence of photoplethysmography has greatly reduced the burden of collecting HR, and research has shown that photoplethysmography can reasonably estimate HR when compared to ECG (87). To date, however, photoplethysmography is much more common in consumer wearable devices than research-grade devices. Thus, research-grade devices that incorporate photoplethysmography are needed.

Out of the sleep studies screened, only 10 studies employed research-grade devices that used MOVE+HR to estimate sleep. There were several studies that included MOVE+HR in addition to one or several other metrics (e.g., eye movement, muscle activity, blood saturation, breathing rate, etc.). However, these studies were excluded because many of these measures are not feasible for use in free-living settings. Moreover, of the 10 studies included, none parceled out the ability of movement or HR alone to predict sleep. It is well established that accelerometry can reliably detect sleep (sensitivity) but has poor detection of wake (specificity) when compared to the gold standard polysomnography (24). This has led to the over or underestimation of total sleep time (28). There is promising evidence that consumer wearable devices incorporating HR via photoplethysmography may improve detection of wake. For instance, in reference to polysomnography, a systematic review and meta-analysis including 4 comparisons (N = 153) reported specificity between 0.58 and 0.69 for recent Fitbit models (29). This is considerably higher than the 0.50 specificity typically achieved by research-grade accelerometers. However, since Fitbits and other consumer wearable devices use proprietary algorithms to predict sleep, it is impossible to know exactly how the addition of HR is influencing the accuracy of their predictions. The studies included in this review all relied upon a combination of research-grade devices and/or custom-made devices (17, 33, 39, 40, 47-49, 52, 74, 83). The lack of a single device that collects both HR and movement likely complicates data collection, increases participant burden, and reduces the use of HR and movement in studies that are attempting to estimate free-living sleep. Thus, there is a clear need for research-grade devices that incorporate HR technology and accelerometry to be developed and validated. One such device is the Actiheart monitor that allows for the prediction of sleep. However, our systematic review of the literature found no studies that have attempted to validate Actiheart's ability to predict sleep metrics.

Strengths: This systematic review and meta-analysis has several strengths. First, the review was guided by the PRISMA guidelines for conducting systematic reviews and meta-analyses, increasing the confidence that the findings accurately reflect the state of the current literature. Second, this meta-analysis quantified estimates of sleep based on MOVE+HR and estimates of physical activity based on movement, HR, and MOVE+HR, extending our knowledge of the most accurate proxy methods to assess physical activity and sleep. Third, an equivalence testing approach was adopted to rigorously test the validity of these metrics for predicting physical activity and sleep.

Limitations: However, the findings of this systematic review and meta-analysis must also be interpreted considering its limitations. First, no included sleep studies reported the ability of movement or HR exclusively to predict sleep. Thus, it was impossible to ascertain if MOVE+HR is superior to movement or HR alone for predicting sleep. This limitation is likely due to the exclusion of consumer wearable devices, as some of these devices incorporate MOVE+HR to estimate sleep. In fact, multiple studies evaluating sleep metrics estimated by multichannel (i.e., movement and HR signals) consumer wearable devices and research-grade accelerometers demonstrate these consumer wearable devices perform as well as, if not better than previously validated research-grade devices, when compared to polysomnography (7, 12, 43, 58). This may suggest that future meta-analyses should consider evaluating both research-grade and consumer wearable devices when evaluating sleep estimates based on movement, HR, or MOVE+HR. Second, studies reported a wide variety of validity metrics (mean absolute percent error, percent error, root mean square error/standard error of estimate, mean absolute error, etc.). In many studies, variability around these metrics (e.g., standard deviation) were not presented, making it impossible to employ traditional meta-analytic techniques for pooling estimates from multiple studies. This did not allow for generation of SMD estimates when metrics were produced for different outcomes (e.g., energy expenditure, physical activity energy expenditure, metabolic equivalents, total sleep time, sleep stage, sleep efficiency, body posture). However, based on recommendations from previous literature (82), we reported mean percent error and mean absolute percent error to meta-analytically evaluate error at the group and individual levels. We also calculated weighted means for proportion of shared variance (i.e., r^2) and percent accuracy since these metrics do not include variability around the point estimates.

Conclusion: For estimates of physical activity, findings revealed no statistically significant differences between MOVE+HR and movement or HR only when compared to criterion indicators of physical activity. A lack of significant equivalence was also observed for estimates of physical activity based on MOVE+HR and movement or HR only compared to criterion indicators. The overall lack of statistical difference and equivalence for all proxy indicators may highlight a limitation of measurement studies in general—estimates of physical activity at the group level are generally closer to criterion estimates, while estimates for any individual in the group may be inaccurate, suggesting there is room to improve upon estimations of physical activity at the individual level in able-bodied human populations. Also, the large heterogeneity across physical activity studies, including different measurement tools, placement of devices, and range of activities may have contributed to the lack of statistical difference and equivalence between the proxy indicators. However, we did not account for heterogeneity in this study. For sleep, all included studies based their estimates solely on MOVE+HR, making it impossible to ascertain whether MOVE+HR is more accurate than either method alone. A lack of statistical equivalence was also observed for estimates of sleep based on MOVE+HR, which may have been due to the variety of sleep measurement tools used (see Table 2) to estimate sleep metrics.

Supplementary Table 1. Search strategy.

Group	Search terms	Search strategy		Search
Outcome	physical activity sedentary sleep nap circadian	#1	TITLE: "Physical Activity" OR ABSTRACT "Physical Activity" TITLE: Sedentary OR ABSTRACT: Sedentary TITLE: Sleep OR ABSTRACT: Sleep TITLE: Nap OR ABSTRACT: Nap TITLE: circadian OR ABSTRACT: circadian	# 1 AND #2
Design	Calib*, Valid*, reliability, accuracy	#2	Calib*OR Valid*OR reliability OR accuracy	

Supplementary Table 2. Exclusion and inclusion terms for pre-screen in EndNote and Filter Terms in Covidence.

Exclusion Terms	Inclusion Terms
Self-report	Actiheart
Psychometric	Multisensor
Scale	Polar
Questionnaire	Sensewear
Cohort	Bland-Altman
Survey	Electrocardiogram
Perception	PAEE
Systematic Review	Heart rate
IPAQ	Heartrate
Rats	Accelerometry
Mouse	Plethysmography
Mice	Sphygmomanometer
Rodent	Oscillometer
Cell	Aortic pressure
Renal	Wrist worn
Antibody	Chest band
Primates	Arterial
Mammal	*Tonometry*
Gene	ECG
CD4	Ballistocardiogram
Monkeys	Motion
Rabbit	Cardio
Pig	Actigraph wGT3X-BT monitors
Horse	Actigraph
Drosophilia	Actipal
Qualitative	Cosmed
Conference	WatchPAT
Dissertation	
Sudden infant death	
Frail	
Abstract	
Poster	

Supplementary Table 3. Quality assessment of the included physical activity studies.

	Physical Activity																												
Item	Corder 2005	Zakeri 2008	Zakeri 2010	Zakeri 2013	Silva 2015	Brage 2004	Brage 2005	Villars 2012	Strath 2001	Eston 1998	Bussman 1998	Ojiambo 2012	Santos 2014	Rumo 2011	Assah 2011	Spierer 2011	Schneller 2015	Rennie 2000	Barreira 2009	Butte 2014	Campbell 2012	Corder 2007	O'Driscoll 2021	Ellis 2014	Garnotel 2018	Gilgen-Ammann 2021	Hernandez-Vicente 2016	Lof 2013	Moon 1996
1																													
2																													
3																													
4																													

1=Appropriate criterion measure of physical activity or sleep is used, 2=A wide variety of activities are included in the protocol (at least 6 activities), 3=Accelerometer data is collected with epochs of less than 60 seconds, 4=More than 10 participants

Supplementary Table 4. Quality assessment of the included sleep studies.

	Sleep									
Item	Yoshihi 2021	Kortelainen 2010	Mendez 2009	Takano 2019	Herscovici 2007	Muzet 2016	Migliorini 2010	DaWoon 2014	Kuula 2021	Mitsukura 2020
1										
2										
3										
4										

1=Appropriate criterion measure of physical activity or sleep is used, 2=A wide variety of activities are included in the protocol (at least 6 activities), 3=Accelerometer data is collected with epochs of less than 60 seconds, 4=More than 10 participants

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences for the University of South Carolina Research Center for Child Well-Being Award Number P20GM130420 and the National Institute of Diabetes and Digestive and Kidney Diseases Award Numbers F31DK136205 & R01DK129215 of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Adenauer H, Catani C, Keil J, Aichinger H, Neuner F. Is freezing an adaptive reaction to threat? Evidence from heart rate reactivity to emotional pictures in victims of war and torture. Psychophysiology 47(2): 315-322, 2010.

2. Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC, Wareham NJ. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in cameroon. Int J Epidemiol 40(1): 112-120, 2011.

3. Barreira TV, Kang M, Caputo JL, Farley RS, Renfrow MS. Validation of the actiheart monitor for the measurement of physical activity. Int J Exerc Sci 2(1): 60-71, 2009.

4. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor actiheart. Eur J Clin Nutr 59(4): 561-570, 2005.

5. Brage S, Brage N, Franks PW, Ekelund U, Wong MY, Andersen LB, Froberg K, Wareham NJ. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol (1985) 96(1): 343-351, 2004.

6. Brewer W, Swanson BT, Ortiz A. Validity of fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 3(1): e000254, 2017.

7. Burkart S, Beets MW, Armstrong B, Hunt ET, Dugger R, von Klinggraeff L, Jones A, Brown DE, Weaver RG. Comparison of multichannel and single-channel wrist-based devices with polysomnography to measure sleep in children and adolescents. J Clin Sleep Med 17(4): 645-652, 2021.

8. Bussmann JB, Tulen JH, van Herel EC, Stam HJ. Quantification of physical activities by means of ambulatory accelerometry: A validation study. Psychophysiology 35(5): 488-496, 1998.

9. Butte NF, Wong WW, Adolph AL, Puyau MR, Vohra FA, Zakeri IF. Validation of cross-sectional time series and multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents using doubly labeled water. J Nutr 140(8): 1516-1523, 2010.

10. Butte NF, Wong WW, Lee JS, Adolph AL, Puyau MR, Zakeri IF. Prediction of energy expenditure and physical activity in preschoolers. Med Sci Sports Exerc 46(6): 1216-1226, 2014.

11. Campbell N, Prapavessis H, Gray C, McGowan E, Rush E, Maddison R. The actiheart in adolescents: A doubly labelled water validation. Pediatr Exerc Sci 24(4): 589-602, 2012.

12. Chee N, Ghorbani S, Golkashani HA, Leong RLF, Ong JL, Chee MWL. Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. Nat Sci Sleep 13: 177-190, 2021.

13. Corder K, Brage S, Mattocks C, Ness A, Riddoch C, Wareham NJ, Ekelund U. Comparison of two methods to assess PAEE during six activities in children. Med Sci Sports Exerc 39(12): 2180-2188, 2007.

14. Corder K, Brage S, Wareham NJ, Ekelund U. Comparison of PAEE from combined and separate heart rate and movement models in children. Med Sci Sports Exerc 37(10): 1761-1767, 2005.

15. Crouter SE, Churilla JR, Bassett DR. Accuracy of the actiheart for the assessment of energy expenditure in adults. Eur J Clin Nutr 62(6): 704-711, 2008.

16. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. Eur J Appl Physiol 98(6): 601-612, 2006.

17. Da Woon J, Hwang SH, Yoon HN, Lee Y-JG, Jeong D-U, Park KS. Nocturnal awakening and sleep efficiency estimation using unobtrusively measured ballistocardiogram. IEEE Trans Biomed Eng 61(1): 131-138, 2014.

18. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, Colrain IM. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 158: 143-149, 2016.

19. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of fitbit charge 2™ compared with polysomnography in adults. Chronobiol Int 35(4): 465-476, 2018.

20. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. Med Sci Sports Exerc 50(4): 837-845, 2018.

21. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 35(11): 2191-2203, 2014.

22. Eston RG, Rowlands AV, Ingledew DK. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. J Appl Physiol 84(1): 362-371, 1998.

23. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. Med Sci Sports Exerc 37(11): S523-S530, 2005.

24. Galland BC, Short MA, Terrill P, Rigney G, Haszard JJ, Coussens S, Foster-Owens M, Biggs SN. Establishing normal values for pediatric nighttime sleep measured by actigraphy: A systematic review and meta-analysis. Sleep 41(4): 1-16, 2018.

25. Garnotel M, Bastian T, Romero-Ugalde H-M, Maire A, Dugas J, Zahariev A, Doron M, Jallon P, Charpentier G, Franc S. Prior automatic posture and activity identification improves physical activity energy expenditure prediction from hip-worn triaxial accelerometry. J Appl Physiol 124(3): 780-790, 2018.

26. Giles GE, Mahoney CR, Brunyé TT, Taylor HA, Kanarek RB. Stress effects on mood, hpa axis, and autonomic response: Comparison of three psychosocial stress paradigms. PloS One 9(12): e113618, 2014.

27. Gilgen-Ammann R, Roos L, Wyss T, Veenstra BJ, Delves SK, Beeler N, Buller MJ, Friedl KE. Validation of ambulatory monitoring devices to measure energy expenditure and heart rate in a military setting. Physiol Meas 42(8): 085008, 2021.

28. Goldstone A, Baker FC, de Zambotti M. Actigraphy in the digital health revolution: Still asleep? Sleep 41(9): 1-2, 2018.

29. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: Systematic review and meta-analysis. J Med Internet Res 21(11): e16273, 2019.

30. Hanneman SK. Design, analysis, and interpretation of method-comparison studies. AACN Adv Crit Care 19(2): 223-234, 2008.

31. Hedberg E. Robumeta: Stata module to perform robust variance estimation in meta-regression with dependent effect size estimates. Statistical Software Components S457219, Boston College Department of Economics, 2011 (revised 2014).

32. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of polar v800 accelerometer. Ann Transl Med 4(15): 278, 2016

33. Herscovici S, Pe'er A, Papyan S, Lavie P. Detecting rem sleep from the finger: An automatic rem sleep algorithm based on peripheral arterial tone (pat) and actigraphy. Physiol Meas 28(2): 129-140, 2006.

34. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four fitbit and jawbone activity monitors with a research-grade actigraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 52(13): 844-850, 2018.

35. Jordan JB, Korgaokar A, Farley RS, Coons JM, Caputo JL. Caffeine supplementation and reactive agility in elite youth soccer players. Pediatr Exerc Sci 26(2): 168-176, 2014.

36. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev 47(4): 206-214, 2019.

37. Kim Y, Beets MW, Welk GJ. Everything you wanted to know about selecting the "right" actigraph accelerometer cut-points for youth, but…: A systematic review. J Sci Med Sport 15(4): 311-321, 2012.

38. Kim Y, Crouter SE, Lee J-M, Dixon PM, Gaesser GA, Welk GJ. Comparisons of prediction equations for estimating energy expenditure in youth. J Sci Med Sport 19(1): 35-40, 2016.

39. Kortelainen JM, Mendez MO, Bianchi AM, Matteucci M, Cerutti S. Sleep staging based on signals acquired through bed sensor. IEEE Trans Inf Technol Biomed 14(3): 776-785, 2010.

40. Kuula L, Pesonen A-K. Heart rate variability and firstbeat method for detecting sleep stages in healthy young adults: Feasibility study. JMIR MHealth and UHealth 9(2): e24704, 2021.

41. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. Front Psychol 4: 863, 2013.

42. Lakens D. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci 8(4): 355-362, 2017.

43. Lee XK, Chee N, Ong JL, Teo TB, van Rijn E, Lo JC, Chee MWL. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. J Clin Sleep Med 15(9): 1337-1346, 2019.

44. Leonard WR. Measuring human energy expenditure: What have we learned from the flex-heart rate method? Am J Hum Biol 15(4): 479-489, 2003.

45. Löf M, Henriksson H, Forsum E. Evaluations of actiheart, ideea® and rt3 monitors for estimating activity energy expenditure in free-living women. J Nutr Sci 2: e31, 2013.

46. Martinez-Gomez D, Ruiz JR, Ortega FB, Casajús JA, Veiga OL, Widhalm K, Manios Y, Béghin L, González-Gross M, Kafatos A. Recommended levels and intensities of physical activity to avoid low-cardiorespiratory fitness in European adolescents: The Helena study. Am J Hum Biol 22(6): 750-756, 2010.

47. Mendez MO, Matteucci M, Cerutti S, Bianchi AM, Kortelainen JM. Automatic detection of sleep macrostructure based on bed sensors. Annu Int Conf IEEE Eng Med Biol Soc 2009: 5555-5558, 2009.

48. Migliorini M, Bianchi AM, Nisticò D, Kortelainen J, Arce-Santana E, Cerutti S, Mendez MO. Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors. Annu Int Conf IEEE Eng Med Biol Soc 2010: 3273-3276, 2010.

49. Mitsukura Y, Sumali B, Nagura M, Fukunaga K, Yasui M. Sleep stage estimation from bed leg ballistocardiogram sensors. Sensors (Basel) 20(19): 5688, 2020.

50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. PLoS Med 6(7): e1000097, 2009.

51. Moon JK, Butte NF. Combined heart rate and activity improve estimates of oxygen consumption and carbon dioxide production rates. J Appl Physiol 81(4): 1754-1761, 1996.

52. Muzet A, Werner S, Fuchs G, Roth T, Saoud JB, Viola AU, Schaffhauser J-Y, Luthringer R. Assessing sleep architecture and continuity measures through the analysis of heart rate and wrist movement recordings in healthy subjects: Comparison with results based on polysomnography. Sleep Med 21: 47-56, 2016.

53. Navalta JW, Stone WJ, Lyons TS. Ethical issues relating to scientific discovery in exercise science. Int J Exerc Sci 12(1): 1-8, 2019.

54. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, Finlayson G, Stubbs J. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 54(6): 332-340, 2020.

55. O'Driscoll R, Turicchi J, Hopkins M, Duarte C, Horgan GW, Finlayson G, Stubbs RJ. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: Validation study. JMIR MHealth and UHealth 9(8): e23938, 2021.

56. Ojiambo R, Konstabel K, Veidebaum T, Reilly J, Verbestel V, Huybrechts I, Sioen I, Casajús JA, Moreno LA, Vicente-Rodriguez G. Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: The idefics validation study. J Appl Physiol 113(10): 1530-1536, 2012.

57. Paruthi S, Brooks LJ, D'Ambrosio C, Hall WA, Kotagal S, Lloyd RM, Malow BA, Maski K, Nichols C, Quan SF. Recommended amount of sleep for pediatric populations: A consensus statement of the American academy of sleep medicine. J Clin Sleep Med 12(6): 785-786, 2016.

58. Pesonen AK, Kuula L. The validity of a new consumer-targeted wrist device in sleep measurement: An overnight comparison against polysomnography in children and adolescents. J Clin Sleep Med 14(4): 585-591, 2018.

59. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. Physical activity guidelines for americans. JAMA 320(19): 2020-2028, 2018.

60. Pulopulos MM, Vanderhasselt M-A, De Raedt R. Association between changes in heart rate variability during the anticipation of a stressful situation and the stress-induced cortisol response. Psychoneuroendocrinology 94: 63-71, 2018.

61. Rennie K, Rowsell T, Jebb SA, Holburn D, Wareham NJ. A combined heart rate and movement sensor: Proof of concept and preliminary testing study. Eur J Clin Nutr 54(5): 409-414, 2000.

62. Rogers JL, Howard KI, Vessey JT. Using significance tests to evaluate equivalence between two experimental groups. Psychol Bull 113(3): 553-565, 1993.

63. Rowlands AV, Eston RG, Ingledew DK. Measurement of physical activity in children with particular reference to the use of heart rate and pedometry. Sports Med 24(4): 258-272, 1997.

64. Rumo M, Amft O, Tröster G, Mäder U. A stepwise validation of a wearable system for estimating energy expenditure in field-based research. Physiol Meas 32(12): 1983-2001, 2011.

65. Rundo JV, Downey R 3rd. Polysomnography. Handb Clin Neurol 160: 381-392, 2019.

66. Santos DA, Silva AM, Matias CN, Magalhães JP, Fields DA, Minderico CS, Ekelund U, Sardinha LB. Validity of a combined heart rate and motion sensor for the measurement of free-living energy expenditure in very active individuals. J Sci Med Sport 17(4): 387-393, 2014.

67. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the Apple watch. J Med Syst 43(7): 195, 2019.

68. Schneller MB, Pedersen MT, Gupta N, Aadahl M, Holtermann A. Validation of five minimally obstructive methods to estimate physical activity energy expenditure in young adults in semi-standardized settings. Sensors (Basel) 15(3): 6133-6151, 2015.

69. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15(6): 657-680, 1987.

70. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, Wheeler MT, Ashley EA. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 7(2): 3, 2017.

71. Silva AM, Santos DA, Matias CN, Judice PB, Magalhães J, Ekelund U, Sardinha LB. Accuracy of a combined heart rate and motion sensor for assessing energy expenditure in free-living adults during a double-blind crossover caffeine trial using doubly labeled water as the reference method. Eur J Clin Nutr 69(1): 20-27, 2015.

72. Spierer DK, Hagins M, Rundle A, Pappas E. A comparison of energy expenditure estimates from the actiheart and actical physical activity monitors during low intensity activities, walking, and jogging. Eur J Appl Physiol 111(4): 659-667, 2011.

73. Strath SJ, Bassett DR, Swartz AM, Thompson DL. Simultaneous heart rate-motion sensor technique to estimate energy expenditure. Med Sci Sports Exerc 33(12): 2118-2123, 2001.

74. Takano M, Ueno A. Noncontact in-bed measurements of physiological and behavioral signals using an integrated fabric-sheet sensing scheme. IEEE J Biomed Health Inform 23(2): 618-630, 2018.

75. Temple JL, Ziegler AM. Gender differences in subjective and physiological responses to caffeine and the role of steroid hormones. J Caffeine Res 1(1): 41-48, 2011.

76. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 48(13): 1019-1023, 2014.

77. Tudor-Locke C, Barreira TV, Schuna JM, Mire EF, Chaput J-P, Fogelholm M, Hu G, Kuriyan R, Kurpad A, Lambert EV. Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the international study of childhood obesity, lifestyle and the environment (iscole). Int J Behav Nutr Phys Act 12: 11, 2015.

78. Tufanaru C, Munn Z, Stephenson M, Aromataris E. Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. Int J Evid Based Healthc 13(3): 196-207, 2015.

79. Van De Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review. J Sleep Res 20(1pt2): 183-200, 2011.

80. Villars C, Bergouignan A, Dugas J, Antoun E, Schoeller DA, Roth H, Maingon A-C, Lefai E, Blanc S, Simon C. Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. J Appl Physiol 113(11): 1763-1771, 2012.

81. Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, Dinges DF, Gangwisch J, Grandner MA. Recommended amount of sleep for a healthy adult: A joint consensus statement of the American academy of sleep medicine and sleep research society. J Clin Sleep Med 11(6): 591-592, 2015.

82. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 51(8): 1767-1780, 2019.

83. Yoshihi M, Okada S, Wang T, Kitajima T, Makikawa M. Estimating sleep stages using a head acceleration sensor. Sensors (Basel) 21(3): 952, 2021.

84. Zakeri I, Adolph AL, Puyau MR, Vohra FA, Butte NF. Application of cross-sectional time series modeling for the prediction of energy expenditure from heart rate and accelerometry. J Appl Physiol 104(6): 1665-1673, 2008.

85. Zakeri IF, Adolph AL, Puyau MR, Vohra FA, Butte NF. Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. J Appl Physiol 108(1): 128-136, 2010.

86. Zakeri IF, Adolph AL, Puyau MR, Vohra FA, Butte NF. Cross-sectional time series and multivariate adaptive regression splines models using accelerometry and heart rate predict energy expenditure of preschoolers. J Nutr 143(1): 114-122, 2013.

87. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of wrist-worn photoplethysmography devices to measure heart rate: A systematic review and meta-analysis. J Sports Sci 38(17): 2021-2034, 2020.