Article:

O'Driscoll, R orcid.org/0000-0003-3995-0073, Turicchi, J orcid.org/0000-0003-1174-813X, Beaulieu, K orcid.org/0000-0001-8926-6953 et al. (5 more authors) (2020) How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. British Journal of Sports Medicine, 54 (6). pp. 332-340. ISSN 0306-3674

https://doi.org/10.1136/bjsports-2018-099643

How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis.

Ruairi O'Driscoll,[1] Jake Turicchi,[1] Kristine Beaulieu,[1] Sarah Scott,[1] Jamie Matu,[2] Kevin Deighton,[3] Graham Finlayson,[1] R. James Stubbs[1]

[1]Appetite Control and Energy Balance Group, School of Psychology, University of Leeds, Leeds, U.K.

[2]Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, U.K.

[3]Institute for Sport, Physical Activity & Leisure, Leeds Beckett University, Leeds, U.K.

Corresponding author:

Ruairi O'Driscoll
Appetite Control and Energy Balance Group
University of Leeds,
Leeds, U.K.
LS2 9JT
Psrod@leeds.ac.uk

Word count:

4493

Abstract

Objective To determine the accuracy of wrist and arm-worn activity monitors' estimates of energy expenditure (EE).

Data sources SportDISCUS (EBSCOHost), PubMed, Medline (Ovid), PsycINFO (EBSCOHost), EMBASE (Ovid) and CINAHL (EBSCOHost).

Design A random effects meta-analysis was performed to evaluate the difference in EE estimates between activity monitors and criterion measurements. Moderator analyses were conducted to determine the benefit of additional sensors and to compare the accuracy of devices used for research purposes with commercially available devices.

Eligibility criteria We included studies validating EE estimates from wrist or arm-worn activity monitors against criterion measures (indirect calorimetry, room calorimeters and doubly labelled water) in healthy adult populations.

Results 60 studies (104 effect sizes) were included in the meta-analysis. Devices showed variable accuracy depending on activity type. Large and significant heterogeneity was observed for many devices (I^2 >75%). Combining heart rate or heat sensing technology with accelerometry decreased the error in most activity types. Research-grade devices were statistically more accurate for comparisons of total EE but less accurate than commercial devices during ambulatory activity and sedentary tasks.

Conclusions EE estimates from wrist and arm-worn devices differ in accuracy depending on activity type. Addition of physiological sensors improves estimates of EE and research-grade devices are superior for total EE. These data highlight the need to improve estimates of EE from wearable devices and one way this can be achieved is with the addition of heart rate to accelerometry.

Registration PROSPERO CRD42018085016.

Keywords*: Energy expenditure, Accelerometer, Meta-analysis, Wrist, Validation.*

Device abbreviations*: Actical (ACT), Actigraph GT3X (AGT3X), Apple watch (AW), Apple Watch series 2 (AWS2), Beurer (BA) Basis b1 (BB1), Bodymedia CORE armband (BMC), Basis Peak (BP), Epson Pulsense (EP), ePulse Personal Fitness Assistant (EPUL), Fitbit Blaze (FB), Fitbit Charge (FC), Fitbit Charge 2 (FC2), Fitbit Charge HR (FCHR), Fitbit Flex (FF), Garmin Forerunner 225 (GF225), Garmin Forerunner 920XT (GF920XT), Garmin Vivoactive (GVA), Garmin Vivofit (GVF), Garmin Vivosmart (GVS), Garmin Vivosmart HR (GVHR), Jawbone UP (JU), Jawbone UP24 (JU24), LifeChek calorie sensor (LC), Mio Alpha (MA), Microsoft band (MB), Misfit Shine (MS), Polar: AW360 (PA360), Nike Fuel band (NF), Polar Loop (PL), Polar: AW200 (PO200), Samsung Gear S (SG), SenseWear Armband (SWA), SenseWear Armband Pro 2 (SWA p2), SenseWear Armband Pro 3 (SWA p3), SenseWear Armband MINI (SWAM), TOMTOM Touch (TT), Vivago (V), Withings Pulse (WP), Withings Pulse O2 (WPO).*

What is already known on this topic?

- Wrist or arm-worn devices incorporating multiple sensors are increasingly common and many devices provide estimates of energy expenditure. It is important to determine their validity overall and in different activity types.
- It is not clear which specific sensors or combinations of sensors provide the most accurate estimates of energy expenditure.
- It is unclear whether research-grade devices are more accurate than commercial devices.

What this study adds

- The accuracy in energy expenditure estimates from activity monitors varies between activities.
- Larger error is observed from devices employing accelerometry alone; the addition of heart rate sensing improves estimates of energy expenditure in most activities.
- In some activity types, research-grade devices are not superior to commercial devices.

Introduction

The prevalence of obesity has tripled in the last 40 years [1] and it has been estimated that by 2050, 60% of males and 50% of females may be obese [2]. Obesity is the result of a chronic imbalance between energy intake (EI) and energy expenditure (EE) [3] driven by physiological, psychological and environmental factors.

Doubly-labelled water (DLW) is considered the gold standard for the measurement of free-living EE [4]; however, the considerable costs and analytical requirements limit its feasibility in large cohort studies [5]. Indirect calorimetry methods represent the most commonly employed criterion measure for assessment of the energy cost of an activity but again are limited to structured activities usually within a laboratory [6]. Wearable activity monitors are increasingly popular for the estimation of EE [7].

Wearable devices which use triaxial accelerometry to derive an estimate of EE have been available for research purposes for some time [8]. These devices are worn on the hip, thigh or lower back, as proximity to the centre of mass more accurately reflects the energy cost of movement [9]; however, participant comfort and compliance is a recognised issue [10] and therefore traditional wear devices have limited long-term, free-living measurement capability. Use of wrist-worn activity monitors by both consumers and researchers has dramatically increased [11] facilitated by improved battery longevity and miniaturization of hardware required to produce interpretable data [12]. Recent consumer devices include triaxial accelerometers, heat sensors and photoplethysmography heart rate sensors [13]. This information can be incorporated to improve the estimation of EE relative to accelerometry alone [14]. However, their accuracy compared with criterion measures is questionable [15] and may vary with the type and intensity of activity [16].

This meta-analysis aimed to investigate the accuracy of EE estimates from current wrist or arm-worn devices during different activities. Given the recent popularity wrist and arm-worn activity monitors, it is critical to determine their validity for the estimation of EE [17]. Secondary aims were to investigate the usefulness of specific sensors within devices, and compare commercial and research-grade devices. We hypothesised that the addition of physiological data to accelerometry within wearable devices will provide a more accurate estimate of EE [18], compared with criterion measures, and that the performance of research-grade devices would be superior to commercial devices.

Methods

This systematic review and meta-analysis adhered to PRISMA diagnostic test accuracy guideline [19] (supplementary material 1) and was prospectively registered in the PROSPERO database (CRD42018085016).

Search strategy

SportDISCUS (EBSCOHost), PubMed, Medline (Ovid), PsycINFO (EBSCOHost), EMBASE (Ovid) and CINAHL (EBSCOHost) were searched for studies published up to 1st December 2017 using terms relevant to the validation of EE estimates from activity monitors against criterion measures with the following strategy ((tracker AND EE) AND validation). The search was updated 15th January 2018. The specific keywords and the full search strategy can be found in supplementary material 2. No language restrictions were applied and in the case of studies available only as an abstract, attempts were made to contact the authors.

Inclusion criteria

We considered laboratory or field validation studies conducted in healthy adults (≥18 years) comparing a criterion measure of EE to an estimate of EE in kilocalories (kcal), kilojoules (kJ) or megajoules (MJ) from an activity monitor. We considered only wrist or arm-worn devices. There is a clear tendency towards wrist worn devices amongst consumer devices and devices worn on alternative anatomical locations produce different accelerometry patterns and therefore estimates of EE [20]. For criterion validation, we considered DLW, indirect calorimetry devices and metabolic chambers [6].

Exclusion criteria

Adults with conditions deemed to produce atypical movement patterns were excluded, including Parkinson's disease, chronic obstructive pulmonary disease, cerebral palsy and amputees. These conditions are often associated with abnormal gait pattern and thus reduce accuracy in EE estimates [21]. Devices requiring external sensors or components were excluded. Studies reporting only accelerometer counts or studies involving post-hoc manipulation of the device output were excluded.

Study selection

Two authors (ROD and JT) independently assessed 100% of titles and abstracts for potential inclusion, with 10% screened independently by a third author (GF). In the case of disagreements between reviewers, the paper was retrieved in full-text and mutual consensus was reached. Remaining articles were screened independently for inclusion at the full-text level by two authors (ROD and JT), with a third author (SS) screening 10%. Similarly, conflicts were resolved by discussion between reviewers.

Data extraction

From each of the included studies, characteristics of participants, validation protocol, criterion measure and the devices tested including model, wear site and output were extracted. Mean difference or EE estimates from the criterion measure and the device were extracted, along with standard deviation (SD), standard error (SE) or 95% confidence intervals (95% CI). If only SE was provided, SE was converted to SD. If data were not provided, authors were contacted to request the raw data. Where values were only presented in figures, a digitiser tool was used [22]. Data was extracted to a specialised spreadsheet and entered into Comprehensive Meta-analysis (CMA) (CMA, version 2; Biostat, Englewood, NJ) for analysis. Data was extracted by one author (ROD) and was cross-checked for data extraction errors. A second author (JT) verified 100% of extracted data and data entered into CMA.

Quality assessment

Risk of bias in included studies was determined using a modified version of the Downs and Black checklist for non-randomised studies [23]. The Downs and Black instrument is an established tool for determination of the quality of a study within a systematic review and meta-analysis [24]. The modified version used in the present study carried a maximum score of 18 and was quantified as: low (≤9, <50%), moderate (>9–14 points, 50–79%), or high (≥15 points, ≥80%) [25]. It contained 17 questions, 10 related to reporting, three to external validity and four to internal validity. The risk of bias assessment was performed independently by two authors (ROD and JT), disagreements were resolved by discussion.

Statistical analysis

Descriptive statistics were calculated for studies included within the meta-analysis. EE estimates from the device and criterion, SD or 95% CI, sample sizes and correlation coefficients for within-activity comparisons for each device were used to calculate effect sizes. Correlation coefficients were based on raw data from previously published studies or were conservatively estimated based on the mean of similar devices (supplementary material 3). Where a study provided data for more than one comparison for one device, the selected outcomes were pooled to provide a single mean and prevent overpowering of a single study. Hedges' g (ES) [26] and 95% CIs were calculated using CMA, in accordance with the majority of studies in the literature testing the mean bias between activity monitors and criterion measures. A negative ES represents an underestimation relative to the criterion and a positive value represents an overestimation. Interpretation of ES was as follows: <0.20 as trivial, 0.20-0.39 as small, 0.40-0.80 as moderate and >0.80 as large [27]. A random effects model was employed for all analyses based on the assumption that heterogeneity would exist between included studies due to the variability in study design [28]. To determine heterogeneity, the I^2 statistic [29] was utilised and >75% was considered to represent large heterogeneity. To determine susceptibility to bias from one study, a leave one out analysis was conducted where the removal of one study would leave at least three studies. The study associated with the greatest change to significance of the effect is reported. To assist interpretation of the error associated with each device, we calculated the percentage error relative for each device using percentage difference and weight within each meta-analysis.

Exploration of small study effects

To examine small study effects, data were visually inspected with funnel plots and subsequently quantified by using Egger's linear regression intercept [30]. A statistically significant Egger's statistic indicates the presence of a small study effect.

Moderators and subgroups

As well as overall, which represents a combination of all subgroups, subgroup meta-analyses were performed for specific activities/categories: 1) activity energy expenditure (AEE) which included comparisons of EE estimates from the device to a criterion during non-specific exercise protocols, circuits, arm ergometer, rowing and resistance exercises; 2) ambulation and stair climbing; 3) cycling; 4) running; 5) sedentary behaviours and household tasks and 6) total energy expenditure (TEE), representing comparisons to DLW.

We conducted moderator analyses by sensors and all devices were grouped based on the inclusion of the following sensor hardware: 1) accelerometry alone (ACC); 2) heart rate alone (HR); 3) accelerometry and heart rate (ACC+HR); 4) accelerometry and heat sensing or galvanic skin response (ACC+HS) and 5) accelerometry, heart rate sensors and heat sensing or galvanic skin response sensors (ACC+HR+HS). Secondly, moderator analyses were conducted by commercial and research-grade devices. Devices produced by Actical, Actigraph and Bodymedia were considered as research-grade and all other devices included in the analysis were considered commercial devices. Comparisons between each moderator employed a random effects model.

Results

Overview

A total of 64 studies were included in the systematic review (Supplementary 4). Four studies could not be synthesised by meta-analysis as mean difference between activity monitors and criterion measurements were not provided [12,31–33]; thus, 60 studies were included in the meta-analysis (figure 1) [10,13,41–50,20,51–60,34,61–70,35,71–80,36,81–88,37–40]. A total of 1946 participants were included, with a mean age of 35 years (range 20 to 86 years). The mean BMI was 24.9 kg/m^2 (range 21.8 to 31.6 kg/m^2). Within the included studies, 104 comparisons between devices and a criterion were included. This represented 58 commercial and 46 research-grade device comparisons. ACC was comprised of 35 comparisons, 1 in HR devices, 20 in ACC+HR devices, 45 in ACC+HS and 3 in ACC+HR+HS. With regard to activity performed, 35 comparisons were classed as AEE, ambulation and stairs included 55 comparisons, 23 were cycling tasks and 38 were running tasks. Sedentary and low-intensity was comprised of 30 comparisons and TEE included 16 comparisons.

Devices

A total of 40 devices were tested in the included studies. One device was forearm-worn, 6 were worn on the upper arm (triceps) and 33 were wrist-worn. Characteristics of the devices, number of studies and weighted percentage error for each device is shown in supplementary materials 5.

Meta-analysis

Individual study effect sizes and allocation to moderator variables are provided in supplementary materials 6. A minimum of three comparisons were required for meta-analysis and as such, we report pooled ES for individual devices or moderators where three or more comparisons were available. Statistical outputs for each device are presented in supplementary materials 7.

Quality assessment

The modified Downs and Black scores revealed a median score of 13, with one study being classed as low quality [69], 48 classed as moderate and 11 classed as high quality (supplementary materials 8). The questions included in the modified tool and percentage of studies fulfilling each question is shown in supplementary materials 9.

Overall

A forest plot of individual devices over all activities is shown in figure 2. Overall, devices underestimated EE (ES: -0.23, 95% CI: -0.44 to -0.04; n=104; p=0.03) and showed significant heterogeneity between devices (I^2 =92.18%; p=<0.001). Significant underestimations relative to criterion measures were observed for the Garmin Vivofit (GVF; ES: -1.09, 95% CI: -1.61 to -0.56; n=5; p<0.001) and the Jawbone UP24 (JU24; ES: -1.16, 95% CI: -1.79 to -0.53; n=3; p<0.001). The SenseWear Armband Pro3 (SWA p3) also underestimated EE (ES: -0.32. 95% CI: -0.62 to -0.01; n=12; p=0.04). Sensitivity analysis revealed that the removal of six comparisons altered the significance of the SWA p3 (p>0.05), the most influential of which decreased the ES to -0.19 (95% CI: -0.50 to 0.11; p=0.21) [81]. The Apple watch (AW) Bodymedia CORE armband (BMC), Fitbit charge HR (FCHR), Fitbit Flex (FF), Jawbone UP (JU), Nike Fuelband (NF), SenseWear Armband (SWA) SenseWear Armband Pro2 (SWA p2), and Mini (SWAM) did not differ significantly from criterion measures. However, sensitivity analysis showed the FCHR differed significantly with the removal of one study (ES: 0.34, 95% CI: 0.20 to 0.49; p<0.001) [88]. The NF was the only device that did not display significant heterogeneity between studies (I^2 =25.44%; p=0.26), with the remaining devices having I^2 values ≥ 66.91% (all p≤0.05). No device showed evidence of small study effects.

AEE

A forest plot of individual devices during activities classed as AEE is shown in supplementary materials 10. For AEE, the pooled estimate of all devices was a non-significant tendency to underestimate EE compared with criterion measures (ES: -0.34, 95% CI: -0.71 to 0.04; n=35; p=0.08) and significant heterogeneity was observed between devices (I^2 =94.94%; p<0.001). The SWA p2 underestimated EE (ES: -0.78, 95% CI: -1.48 to -0.08; n=3; p=0.03) and had moderate, non-significant heterogeneity (I^2 =64.19%; p=0.06). The BMC, NF, SWA and SWAM did not differ significantly from criterion measures but all displayed significant heterogeneity. No device showed evidence of small study effects.

Ambulation and stairs

A forest plot of individual devices during ambulation and stair climbing is shown in figure 3. The pooled estimate of all devices did not differ from criterion measures (ES: -0.09, 95% CI: -0.45 to 0.27; n=55; p=0.62) and significant heterogeneity was observed between devices (I^2 =93.74%; p<0.01). The FCHR (ES: 0.78, 95% CI: 0.27 to 1.29; n=5; p=0.002) and FF (ES: 1.10, 95% CI: 0.43 to 1.77; n=3; p=0.001) overestimated EE. The GVF underestimated EE (ES: -1.24, 95% CI: -1.86 to -0.62; n=4; p<0.01), however, sensitivity analysis revealed that the removal of two comparisons significantly altered the mean effect (p>0.05) the most influential significantly altered the mean effect to ES: -1.32 (95% CI: -2.73 to 0.08; p=0.07) [34]. Further, there was evidence of small study effects (intercept= -13.76, 95% CI: -19.72 to -7.80; p=0.01). The SWA overestimated EE (ES: 0.79, 95% CI: 0.25 to 1.33; n=5; p<0.01) and sensitivity analysis revealed that the removal of four comparisons significantly altered the mean effect (p>0.05) the most influential significantly altered the mean effect to ES: 0.33 (95% CI: -0.26 to 0.92; p=0.28) [56]. The AW, JU, SWA p3 and SWAM did not differ significantly from criterion measures. The mean effect of the SWAM was significantly altered by the removal of two studies; the removal of the most influential study yielded a significant overestimation (ES: 0.57, 95% CI: 0.20 to 0.94; p=0.003) [87]. All devices showed significant heterogeneity.

Cycling

A forest plot of individual devices during cycling is shown in supplementary materials 10. The pooled estimate of all devices was significantly lower than criterion measures (ES: -0.73, 95% CI: -1.39 to -0.06; n=23; p=0.03) and significant heterogeneity was observed between devices (I^2 =94.74%; p<0.01). The SWA did not differ significantly from criterion but showed significant heterogeneity (I^2 =89.39%; p<0.001). The SWA p3 did not differ from criterion measures and showed moderate heterogeneity (I^2 =54.95%; p=0.11).

Running

A forest plot of individual devices during running is shown in supplementary materials 10. The pooled estimate was not statistically different from criterion measures (ES: -0.08, 95% CI: -0.41 to 0.25; n=38; p=0.65) and significant heterogeneity was observed between devices (I^2 =92.05%; p=<0.001). The FCHR, GVF and SWA did not differ from criterion measures.

Sensitivity analysis revealed the removal of one study changed the overall effect for the FCHR (ES: 0.59, 95% CI: 0.28 to 0.90; p<0.001) [87]. Significant heterogeneity was observed for the FCHR (I^2 =66.8%; p=0.03) and SWA (I^2 =96.79; p<0.001), but not for the GVF (I^2 =46.39%; p=0.15).

Sedentary and household tasks

A forest plot of individual devices during sedentary and household tasks is shown in figure 4. The pooled effect was not statistically different from criterion measures (ES: -0.09, 95% CI: -0.51 to 0.32; n=30; p=0.66) and significant heterogeneity was observed between devices (I^2 =94.84%; p<0.001). The AW, FCHR and SWAM were not statistically different from criterion measures. The SWA p3 overestimated EE (ES: 0.67, 95% CI: 0.00 to 1.34; p=0.049). Sensitivity analysis revealed that the removal of three studies changed the mean effect, the most influential of which decreased the ES to 0.41 (95% CI: -0.01 to 0.82; p=0.05) [42]. Observed heterogeneity was significant for the AW, SWA p3 and SWAM. The FCHR had moderate, non-significant heterogeneity (I^2 =59.60%; p=0.60).

TEE

A forest plot of individual devices for the measurement of TEE is shown in figure 5. The pooled effect for TEE showed a significant underestimation of EE (ES: -0.68, 95% CI: -1.15 to -0.21; n=16; p= p=0.005) and significant heterogeneity was observed between devices (I^2 =92.17%; p<0.01). The SWA p3 did not differ significantly from criterion measures and showed significant heterogeneity (I^2 =94.20%; p=0.001).

Moderator analyses

The results of moderator analyses are shown in table 1. Overall, there was a significant difference between sensors (p=0.003). Pooled estimate of EE from ACC+HR and ACC+HS was not statistically different from criterion but ACC+HS showed a non-significant tendency for underestimation, and ACC and ACC+HR+HS both significantly underestimated EE. In the AEE comparison, there was no statistical difference between sensors, but ACC+HS significantly underestimated EE, ACC showed a non-significant tendency for underestimation and ACC+HR did not differ significantly from criterion measures. During ambulation and stair climbing, a significant difference between sensors was observed, with estimates of EE from ACC+HR and ACC+HS being significantly higher than criterion. In cycling, significant differences were observed between sensors, with ACC devices underestimating EE. During running activities, none of the pooled mean estimates were significantly different from criterion. For sedentary and household tasks, a significant difference was observed between sensors; ACC+HR was not different from criterion measures whereas ACC and ACC+HS underestimated and overestimated EE respectively. For TEE, sensors differed significantly; ACC underestimated EE, whereas ACC+HS did not differ significantly from criterion.

When analysed by commercial and research-grade devices, no significant difference was observed overall, for AEE, cycling or running. For both the ambulation and stairs comparison and the sedentary and household tasks comparison, commercial devices were closer to criterion measurements, with research grade devices significantly overestimating. For TEE, research-grade devices were superior, with commercial devices significantly underestimating EE.

Moderator variable	Subgroup level	*p*-value	Hedges' g (95% CI)
Overall activities			
Sensors	ACC (n=35)	<0.01	-0.36 (-0.55, -0.17)*
	ACC + HR (n=20)		0.06 (-0.18, 0.31)
	ACC + HR + HS (n=3)		-0.99 (-1.65, -0.33)*
	ACC + HS (n=45)		-0.151 (-0.32, 0.01)
Device grade	Commercial (n=58)	0.27	-0.269(-0.42, -0.12)*
	Research (n=46)		-0.141 (-0.31, 0.03)
AEE			
Sensors	ACC (n=8)	0.19	-0.40 (-0.84, 0.04)
	ACC + HR (n=9)		-0.04 (-0.47, 0.38)
	ACC + HS (n=16)		-0.32 (-0.63, -0.01)*
Device grade	Commercial (n=18)	0.62	-0.38 (-0.67, -0.08)*
	Research (n=17)		-0.27 (-0.57, 0.04)
Ambulation and stairs			
Sensors	ACC (n=24)	0.01	-0.23 (-0.51, 0.06)
	ACC + HR (n=10)		0.45 (0.02, 0.87)*
	ACC + HS (n=19)		0.40 (0.08, 0.72)*
Device grade	Commercial (n=35)	0.05	-0.04 (-0.28, 0.20)
	Research (n=20)		0.37 (0.05, 0.68)*
Cycling			
Sensors	ACC (n=3)	<0.01	-3.75 (-4.65, -2.85)*
	ACC + HR (n=9)		-0.04 (-0.47, 0.40)
	ACC + HS (n=9)		-0.41 (-0.84, 0.02)
Device grade	Commercial (n=14)	0.28	-0.82 (-1.30, -0.35)*
	Research (n=9)		-0.41 (-0.99, 0.17)
Running			
Sensors	ACC (n=19)	0.18	-0.06 (-0.364, 0.24)
	ACC + HR (n=7)		0.34 (-0.15, 0.82)
	ACC + HS (n=10)		-0.36 (-0.78, 0.05)
Device grade	Commercial (n=28)	0.08	0.06 (-0.18, 0.30)
	Research (n=10)		-0.36 (-0.76, 0.04)
Sedentary and household			
Sensors	ACC (n=6)	<0.01	-0.65 (-1.16, -0.13)*
	ACC + HR (n=9)		0.14 (-0.28, 0.57)
	ACC + HS (n=13)		0.41 (0.06, 0.75)*
Device grade	Commercial (n=17)	<0.01	-0.27 (-0.59, 0.05)
	Research (n=13)		0.41 (0.05, 0.77)*
TEE (DLW)			
Sensors	ACC (n=5)	<0.01	-1.24(-1.66, -0.81)*
	ACC + HS (n=10)		-0.13(-0.397, 0.32)
Device grade	Commercial (n=6)	<0.01	-1.13(-1.51, -0.76)*
	Research (n=10)		-0.13 (-0.39, 0.14)

Discussion

Given the clinical and consumer uptake of wrist and arm-worn activity monitors which can be used for the estimation of EE, the aims of this meta-analysis were (i) to determine the relative accuracy of current devices, (ii) to investigate the importance of specific sensors within devices and (iii) to compare commercial and research-grade devices.

For devices with sufficient comparisons to be analysed separately from the main pooled effect, significant error relative to criterion measures was observed for Garmin, Fitbit, Jawbone and Bodymedia products. Garmin, Fitbit and Jawbone represent a major share of the commercial wearable market [73] and Bodymedia products are widely used in research and have been since 2004 [59]. Whilst it is initially encouraging that the ES for many devices was not significantly different from criterion, the 95% CI observed in many cases indicates the potential for these devices to produce erroneous estimates of mean EE and as such we would be hesitant to consider any device sufficiently accurate. A 10% ‘equivalence zone’ has been suggested previously [65] and with the exception of the Nike Fuel band, in which all three studies reported a mean error <10% [65,79,82], no device pooled in this meta-analysis consistently met this criteria. The SenseWear armband Mini was the most accurate device overall but error reported in studies ranged from -21.27% [87] to 14.76% [39]. Studies in this analysis followed the manufacturer’s instructions for setup, with researchers ensuring the position of the device and characteristics such as height, weight, sex and age were correct. In free-living environments the lack of researcher presence could yield greater error than observed in this analysis [17], as indicated by the moderate, significant underestimation for the pooled effect in the TEE subgroup.

An accurate yet affordable measure of TEE, with a measure of change in energy storage, could theoretically be used to retrospectively determine free-living EI in large cohorts [89]. In this context, TEE may be considered the most important activity subgroup in this meta-analysis, however, the most variable and unpredictable component of TEE is EE during activity [6]. In agreement with previous studies [13,45,52], we have shown that the accuracy of devices differs by activity and this may be related to the inability of devices to differentiate between activity types. For a device to accurately estimate TEE between individuals, it must accurately estimate the energy cost of a wide range of activities however, some activities may require greater focus. The majority of EE is attributable to rest or non-exercise activity [6] so error here could have a great impact on the error in TEE. The Fitbit Charge HR was the most tested commercial device in this analysis and it showed a trivial, non-significant ES overall and during sedentary tasks but a moderate to large and significant overestimation during ambulatory activity. Considering that ambulatory activity is central to public health guidelines worldwide [90], the implications of this finding may be great for estimates of TEE.

The observed error for different activity types may be because current algorithms do not take physical activity type or bodily posture into account [91]. Indeed, activity recognition is considered an important direction for wearable technology [11] and has been used to improve estimates of EE [92]. Montoye et al have shown that accelerometers worn on the wrists and thigh can be used to predict activity type [93]. The SenseWear software employs complex pattern-recognition algorithms to determine activity type [45] which likely contributed to the trivial or small ES observed for the SenseWear Armband Mini in all comparisons. The challenges associated with activity recognition have been reviewed recently [94] and as this technology develops, activity-specific EE prediction equations may offer the opportunity to reduced errors associated with activity types.

Sensors

A 2012 review concluded that multisensory and triaxial accelerometry devices improve estimates of EE, relative to uniaxial devices [21]. Due to recent technological advancements, triaxial accelerometry, as well as heart rate or heat sensing technology are commonplace in newer devices [48]. We hypothesised that the addition of this technology to accelerometry would improve estimates of EE. Overall, this meta-analysis shows that the inclusion of heart rate or heat sensors in devices can improve estimates of EE relative to accelerometry alone. Indeed, it is established that accelerometry is limited for non-weight-bearing activities [84], and accelerometry underestimated EE during cycling activities in our analysis. Significant underestimations were also observed during sedentary and household tasks and TEE, which is likely a product of the limited arm movements associated with these activities.

Accelerometry and heart rate devices moderately overestimated EE during ambulation and stair climbing. Some of this error may be attributable to the individual variability in the relationship between heart rate and EE. Individual calibration of this relationship in the Actiheart device is associated with improved estimates of EE [95] and may offer a means for further reducing the error observed in wrist and arm-worn devices. An alternative explanation for this is the variability in estimates of heart rate from photoplethysmography heart rate sensors. A recent study reported a small mean error of -5.9 bpm in the Fitbit Charge 2, but wide limits of agreement of -28.5 to 16.8 bpm [96] and this variability is a common finding [35,40].

Device Grade

The third aim of this meta-analysis was to compare commercial and research-grade devices. Commercial devices may be developed with affordability and comfort as a primary focus, and as a consequence it may be unreasonable to expect commercial devices to match the validity of research-grade devices. Recent consumer monitors share similar technology with established research-grade multi-sensor devices [48] and this is partially reflected in our results. A benefit of research-grade devices for TEE was observed, but commercial devices were statistically superior in ambulation and during sedentary tasks. Our results question the use of wrist or arm-worn research-grade devices for the validation of newer devices. Comparisons to criterion measures such as DLW or indirect calorimetry are more appropriate when absolute accuracy is required [6]. Further, it is important to highlight that other research-grade devices, for instance the Actiheart, which is worn on the chest [95], are likely to be more accurate than research-grade devices included in this study [48]. Further research is needed to establish whether research-grade devices that are worn in other locations such as the chest, hip or thigh outperform consumer based devices.

Limitations

Separate pooled analyses to determine the accuracy of individual activity monitors were performed for a limited number of devices due to the small number of comparisons available for the remaining devices (i.e., less than three comparisons). This limitation is inevitable considering the large number of activity monitors included in this review. Nevertheless, the inclusion of all devices in the overall pooled analysis provides an extensive and robust evaluation of the difference in EE outcomes between activity monitors and criterion measures.

The majority of analyses conducted within this review demonstrated large heterogeneity within and between devices which remained after moderating by specific devices and activity. Such heterogeneity is not unexpected and in many cases may be attributable to disparity in the protocols employed [97]. Indirect calorimetry systems were the most commonly used criterion measure but EE estimates may differ by up to 5.2% depending on the equations used [98]. EE is likely to be elevated in the period following higher intensity exercise and the inclusion of only the steady state period may influence the extent to which devices differ from criterion measures [56]. There is also the possibility that the discrepancy between device estimates relates to populations studied [16] for example, a higher BMI [35,40] or age related changes in movement patterns [69]. As few devices currently provide open-access to EE algorithms, the potential for this to create heterogeneity remains uncertain. Despite this, the statistically significant outcomes in many cases suggests a consistent direction in effect sizes for many comparisons and the differences in statistical outcomes between devices are supported by the magnitude of effect sizes.

External validity was low in 46 studies pooled in this meta-analysis, which must be considered when interpreting the present results. It must also be noted that the present analysis was limited to healthy individuals and therefore our results cannot be generalized to populations with conditions that produce abnormal gait patterns.

Lastly, there is a lag between product release and testing in research environments [40] and some of the devices included in this meta-analysis are no longer in production so the continued validation of newer devices is imperative.

Conclusion

This meta-analysis collated studies evaluating the validity of EE estimates by wrist or arm-worn devices. Devices vary in accuracy depending on activity type and the significant heterogeneity means caution must be exercised when interpreting these results. Devices with heart rate sensors often produced better estimates than devices using accelerometry only; however, this was not consistent across all activities. Wrist and arm-worn research-grade devices were more accurate than commercial devices for estimates of TEE but researchers should be aware that such devices do not guarantee superior accuracy. Future research should aim to understand and reduce the error in EE estimates from wrist or arm-worn devices in different activity types. This may be achieved through activity recognition techniques, incorporating physiological measures and exploring the potential for individual calibration of these relationships.

Funding

The research was funded by a University of Leeds PhD studentship. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicting interests

None

Reference list

1 Ells LJ, Demaio A, Farpour-Lambert N. Diet, genes, and obesity. *BMJ* 2018;**360**:k7. doi:10.1136/BMJ.K7

2 Agha M, Agha R. The rising prevalence of obesity. *Int J Surg Oncol* 2017;**2**:e17. doi:10.1097/IJ9.0000000000000017

3 Carneiro IP, Elliott SA, Siervo M, *et al.* Is Obesity Associated with Altered Energy Expenditure? *Adv Nutr* 2016;**7**:476–87. doi:10.3945/an.115.008755.findings

4 Seale JL, Conway JM, Canary JJ. Seven-day validation of doubly labeled water method using indirect room calorimetry. *J Appl Physiol* 1993;**74**:402–9. doi:10.1152/jappl.1993.74.1.402

5 Delany JP. Measurement of energy expenditure. Pediatr. Blood Cancer. 2012;**58**:129–34. doi:10.1002/pbc.23369

6 Hills AP, Mokhtar N, Byrne NM. Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures. *Front Nutr* 2014;**1**:1–16. doi:10.3389/fnut.2014.00005

7 Dhurandhar N V, Schoeller D, Brown AW, *et al.* Energy Balance Measurement: When Something is Not Better than Nothing. *Int J Obes* 2015;**39**:1109–13. doi:10.1038/ijo.2014.199

8 Lyden K, Kozey SL, Staudenmeyer JW, *et al.* A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol* 2011;**111**:187–201. doi:10.1007/s00421-010-1639-8

9 Chen KY, Acra SA, Majchrzak K, *et al.* Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes Technol Ther* 2003;**5**:1023–33. doi:10.1089/152091503322641088

10 Diaz KM, Krupka DJ, Chang MJ, *et al.* Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. *BMC Res Notes* 2016;**9**:213. doi:10.1186/s13104-016-2020-8

11 Wright SP, Hall Brown TS, Collier SR, *et al.* How consumer physical activity monitors could transform human physiology research. *Am J Physiol - Regul Integr Comp Physiol* 2017;**312**:R358–67. doi:10.1152/ajpregu.00349.2016

12 Shcherbina A, Mikael Mattsson C, Waggott D, *et al.* Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med* 2017;**7**:1–12. doi:10.3390/jpm7020003

13 Woodman JA, CROUTER SE, BASSETT DR, *et al.* Accuracy of Consumer Monitors for Estimating Energy Expenditure and Activity Type. *Med Sci Sports Exerc* 2017;**49**:371–7. doi:10.1249/MSS.0000000000001090

14 Silva AM, Santos DA, Matias CN, *et al.* 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* 2015;**69**:20–7. doi:10.1038/ejcn.2014.51

15 Stahl SE, An H-S, Dinkel DM, *et al.* How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc Med* 2016;**2**:e000106. doi:10.1136/bmjsem-2015-000106

16 Koehler K, Drenowatz C. Monitoring energy expenditure using a multi-sensor device-Applications and limitations of the sense wear armband in athletic populations. *Front Physiol* 2017;**8**:1–7. doi:10.3389/fphys.2017.00983

17 Evenson KR, Goto MM, Furberg RD, *et al.* Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;**12**:159. doi:10.1186/s12966-015-0314-1

18 Brage SS, Westgate K, Franks PW, *et al.* Estimation of free-living energy expenditure by heart rate and movement sensing: A doubly-labelled water study. *PLoS One* 2015;**10**:e0137206. doi:10.1371/journal.pone.0137206

19 McInnes MDF, Moher D, Thombs BD, *et al.* Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies. *JAMA* 2018;**319**:388. doi:10.1001/jama.2017.19163

20 Nelson MB, Kaminsky LA, Dickin DC, *et al.* Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. *Med Sci Sports Exerc* 2016;**48**:1619–28. doi:10.1249/MSS.0000000000000933

21 Van Remoortel H, Giavedoni S, Raste Y, *et al.* Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act* 2012;**9**:84. doi:10.1186/1479-5868-9-84

22 Rohatgi. WebPlotDigitizer - Extract data from plots, images, and maps. 2017.https://automeris.io/WebPlotDigitizer/ (accessed 6 Mar 2018).

23 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;**52**:377–84. doi:10.1136/JECH.52.6.377

24 Deeks JJ, Dinnes J, D’Amico R, *et al.* Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**:iii–x, 1-173.http://www.ncbi.nlm.nih.gov/pubmed/14499048 (accessed 23 Feb 2018).

25 MacDonald H V, Johnson BT, Huedo-Medina TB, *et al.* Dynamic Resistance Training as Stand-Alone Antihypertensive Lifestyle Therapy: A Meta-Analysis. *J Am Heart Assoc* 2016;**5**. doi:10.1161/JAHA.116.003231

26 Hedges L V. Distribution Theory for Glass’s Estimator of Effect Size and Related Estimators. *J Educ Stat* 1981;**6**:107. doi:10.2307/1164588

27 Cohen J. *Statistical power analysis for the behavioral sciences*. Academic Press 1977. https://www.sciencedirect.com/science/book/9780121790608 (accessed 23 Feb 2018).

28 Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;**172**:137–59. doi:10.1111/j.1467-985X.2008.00552.x

29 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. doi:10.1002/sim.1186

30 Egger M, Smith GD, Schneider M, *et al.* Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34. doi:10.1136/bmj.315.7109.629

31 Lopez GA, Brønd JC, Andersen LB, *et al.* Validation of SenseWear Armband in children, adolescents, and adults. *Scand J Med Sci Sport* 2018;**28**:487–95. doi:10.1111/sms.12920

32 Reeve MD, Pumpa KL, Ball N. Accuracy of the SenseWear Armband Mini and the BodyMedia FIT in resistance training. *J Sci Med Sport* 2014;**17**:630–4. doi:10.1016/j.jsams.2013.08.007

33 Machač S, Procházka M, Radvanský J, *et al.* Validation of Physical Activity Monitors in Individuals with Diabetes: Energy Expenditure Estimation by the Multisensor SenseWear Armband Pro3 and the Step Counter Omron HJ-720 Against Indirect Calorimetry During Walking. *Diabetes Technol Ther* 2013;**15**:413–8. doi:10.1089/dia.2012.0235

34 Alsubheen SA, George AM, Baker A, *et al.* Accuracy of the vivofit activity tracker. *J Med Eng Technol* 2016;**40**:298–306. doi:10.1080/03091902.2016.1193238

35 Bai Y, Hibbing P, Mantis C, *et al.* Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. *J Sports Sci* 2018;**36**:1734–41. doi:10.1080/02640414.2017.1412235

36 Benito PJ, Neiva C, González-Quijano PS, *et al.* Validation of the SenseWear armband

in circuit resistance training with different loads. *Eur J Appl Physiol* 2012;**112**:3155–9. doi:10.1007/s00421-011-2269-5

37 Berntsen S, Hageberg R, Aandstad A, *et al.* Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med* 2010;**44**:657–64. doi:10.1136/bjsm.2008.048868

38 Berntsen S, Stafne SN, Mørkved S. Physical activity monitor for recording energy expenditure in pregnancy. *Acta Obstet Gynecol Scand* 2011;**90**:903–7. doi:10.1111/j.1600-0412.2011.01172.x

39 Bhammar DM, Sawyer BJ, Tucker WJ, *et al.* Validity of SenseWear(R) Armband v5.2 and v2.2 for estimating energy expenditure. *J Sports Sci* 2016;**34**:1830–8. doi:10.1080/02640414.2016.1140220

40 Boudreaux BD, Hebert EP, Hollander DB, *et al.* Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. *Med Sci Sports Exerc* 2018;**50**:624–33. doi:10.1249/MSS.0000000000001471

41 Brazeau AS, Karelis AD, Mignault D, *et al.* Accuracy of the SenseWear Armband??? during ergocycling. *Int J Sports Med* 2011;**32**:761–4. doi:10.1055/s-0031-1279768

42 Brazeau AS, Suppere C, Strychar I, *et al.* Accuracy of energy expenditure estimation by activity monitors differs with ethnicity. *Int J Sports Med* 2014;**35**:847–50. doi:10.1055/s-0034-1371837

43 Brazeau AS, Beaudoin N, Bélisle V, *et al.* Validation and reliability of two activity monitors for energy expenditure assessment. *J Sci Med Sport* 2016;**19**:46–50. doi:10.1016/j.jsams.2014.11.001

44 Brugniaux J V., Niva A, Pulkkinen I, *et al.* Polar activity watch 200: A new device to accurately assess energy expenditure. *Br J Sports Med* 2010;**44**:245–9. doi:10.1136/bjsm.2007.045575

45 Calabro MA, Lee JM, Saint-Maurice PF, *et al.* Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act* 2014;**11**:119. doi:10.1186/s12966-014-0119-7

46 Calabro MA, Kim Y, Franke WD, *et al.* Objective and subjective measurement of energy expenditure in older adults: A doubly labeled water study. *Eur J Clin Nutr* 2015;**69**:850–5. doi:10.1038/ejcn.2014.241

47 Casiraghi F, Lertwattanarak R, Luzi L, *et al.* Energy Expenditure Evaluation in Humans and Non-Human Primates by SenseWear Armband. Validation of Energy Expenditure Evaluation by SenseWear Armband by Direct Comparison with Indirect

Calorimetry. *PLoS One* 2013;**8**:e73651. doi:10.1371/journal.pone.0073651

48 Chowdhury EA, Western MJ, Nightingale TE, *et al.* Assessment of laboratory and daily energy expenditure estimates from consumer multisensor physical activity monitors. *PLoS One* 2017;**12**:e0171720. doi:10.1371/journal.pone.0171720

49 Colbert LH, Matthews CE, Havighurst TC, *et al.* Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc* 2011;**43**:867–76. doi:10.1249/MSS.0b013e3181fc7162

50 Correa JB, Apolzan JW, Shepard DN, *et al.* Evaluation of the ability of three physical activity monitors to predict weight change and estimate energy expenditure. *Appl Physiol Nutr Metab* 2016;**41**:758–66. doi:10.1139/apnm-2015-0461

51 Diaz KM, Krupka DJ, Chang MJ, *et al.* Fitbit®: an Accurate and Reliable Device for Wireless Physical Activity Tracking. *Int J Cardiol* 2016;**185**:138–40. doi:10.1016/j.ijcard.2015.03.038.FITBIT

52 Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *J Med Eng Technol* 2016;**40**:325–9. doi:10.1080/03091902.2016.1197978

53 Erdogan A, Cetin C, Karatosun H, *et al.* Accuracy of the Polar S810i(TM) Heart Rate Monitor and the Sensewear Pro Armband(TM) to Estimate Energy Expenditure of Indoor Rowing Exercise in Overweight and Obese Individuals. *J Sports Sci Med* 2010;**9**:508–16.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3761702&tool=pmcentrez&rendertype=abstract

54 Fruin ML, Rankin JW. Validity of a multi-sensor Armband in estimating rest and exercise energy expenditure. *Med Sci Sports Exerc* 2004;**36**:1063–9. doi:10.1249/01.MSS.0000128144.91337.38

55 Furlanetto KC, Bisca GW, Oldemberg N, *et al.* Step Counting and Energy Expenditure Estimation in Patients With Chronic Obstructive Pulmonary Disease and Healthy Elderly: Accuracy of 2 Motion Sensors. *Arch Phys Med Rehabil* 2010;**91**:261–7. doi:10.1016/j.apmr.2009.10.024

56 Gastin PB, Cayzer C, Dwyer D, *et al.* Validity of the ActiGraph GT3X+ and BodyMedia SenseWear Armband to estimate energy expenditure during physical activity and sport. *J Sci Med Sport* Published Online First: July 2017. doi:10.1016/j.jsams.2017.07.022

57 Heiermann S, Khalaj Hedayati K, Müller MJ, *et al.* Accuracy of a portable multisensor

body monitor for predicting resting energy expenditure in older people: A comparison with indirect calorimetry. *Gerontology* 2011;**57**:473–9. doi:10.1159/000322109

58 Imboden MT, Nelson MB, Kaminsky LA, *et al.* 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* 2017;:bjsports-2016-096990. doi:10.1136/bjsports-2016-096990

59 Jakicic JM, Marcus M, Gallagher KI, *et al.* Evaluation of the SenseWear Pro Armband to Assess Energy Expenditure during Exercise. *Med Sci Sport Exerc* 2004;**36**:897–904. doi:10.1249/01.MSS.0000126805.32659.43

60 Johannsen DL, Calabro MA, Stewart J, *et al.* Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Med Sci Sports Exerc* 2010;**42**:2134–40. doi:10.1249/MSS.0b013e3181e0b3ff

61 Kim Y, Welk GJ. Criterion validity of competing accelerometry-based activity monitoring devices. *Med Sci Sports Exerc* 2015;**47**:2456–63. doi:10.1249/MSS.0000000000000691

62 King GA, Torres N, Potter C, *et al.* Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med Sci Sports Exerc* 2004;**36**:1244–51. doi:10.1249/01.MSS.0000132379.09364.F8

63 Koehler K, Braun H, De MarIes M, *et al.* Assessing energy expenditure in male endurance athletes: Validity of the sensewear armband. *Med Sci Sports Exerc* 2011;**43**:1328–33. doi:10.1249/MSS.0b013e31820750f5

64 Lee CM, Gorelick M, Mendoza A. Accuracy of an infrared led device to measure heart rate and energy expenditure during rest and exercise. *J Sports Sci* 2011;**29**:1645–53. doi:10.1080/02640414.2011.609899

65 Lee J-MM, Kim Y-WY, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc* 2014;**46**:1840–8. doi:10.1249/MSS.0000000000000287

66 MacKey DC, Manini TM, Schoeller DA, *et al.* Validation of an armband to measure daily energy expenditure in older adults. *Journals Gerontol - Ser A Biol Sci Med Sci* 2011;**66 A**:1108–13. doi:10.1093/gerona/glr101

67 Martien S, Seghers J, Boen F, *et al.* Energy expenditure in institutionalized older adults: Validation of SenseWear Mini. *Med Sci Sports Exerc* 2015;**47**:1265–71. doi:10.1249/MSS.0000000000000529

68 McMinn D, Rowe DA, Murtagh S, *et al.* The effect of a school-based active

commuting intervention on children's commuting physical activity and daily physical activity. *Prev Med (Baltim)* 2012;**54**:316–8. doi:10.1016/j.ypmed.2012.02.013

69 Melanson EL, Dykstra JC, Szuminsky N. A novel approach for measuring energy expenditure in free-living humans. *2009 Annu Int Conf IEEE Eng Med Biol Soc* 2009;**2009**:6873–7. doi:10.1109/IEMBS.2009.5333124

70 Montoye AHK, Mitrzyk JR, Molesky MJ. Comparative Accuracy of a Wrist-Worn Activity Tracker and a Smart Shirt for Physical Activity Assessment. *Meas Phys Educ Exerc Sci* 2017;**21**:201–11. doi:10.1080/1091367X.2017.1331166

71 Murakami H, Kawakami R, Nakae S, *et al.* Accuracy of Wearable Devices for Estimating Total Energy Expenditure. *JAMA Intern Med* 2016;**176**:702. doi:10.1001/jamainternmed.2016.0152

72 Papazoglou D, Augello G, Tagliaferri M, *et al.* Evaluation of a multisensor armband in estimating energy expenditure in obese individuals. *Obesity* 2006;**14**:2217–23. doi:10.1038/oby.2006.260

73 Price K, Bird SR, Lythgo N, *et al.* Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J Med Eng Technol* 2017;**41**:208–15. doi:10.1080/03091902.2016.1253795

74 Reece JD, Barry V, Fuller DK, *et al.* Validation of the SenseWear Armband as a Measure of Sedentary Behavior and Light Activity. *J Phys Act Heal* 2015;**12**:1229–37. doi:10.1123/jpah.2014-0136

75 Rousset S, Fardet A, Lacomme P, *et al.* Comparison of total energy expenditure assessed by two devices in controlled and free-living conditions. *Eur J Sport Sci* 2015;**15**:391–9. doi:10.1080/17461391.2014.949309

76 Ryan J, Gormley J. Measurement of energy expenditure by activity monitors. *Phys Ther Rev* 2013;**18**:239–62. doi:10.1179/1743288X13Y.0000000063

77 Slinde F, Bertz F, Winkvist A, *et al.* Energy expenditure by multisensor armband in overweight and obese lactating women validated by doubly labeled water. *Obesity* 2013;**21**:2231–5. doi:10.1002/oby.20363

78 Smith KM, Lanningham-Foster LM, Welk GJ, *et al.* Validity of the SenseWear® armband to predict energy expenditure in pregnant women. *Med Sci Sports Exerc* 2012;**44**:2001–8. doi:10.1249/MSS.0b013e31825ce76f

79 Stackpool C, Porcari JP, Mikat RP GC and FC. The accuracy of various activity trackers in estimating steps taken and energy expenditure. *J Fit Res* 2013;**53**:1689–99.

doi:10.1017/CBO9781107415324.004

80 St-onge M, Mignault D, Allison DB, *et al.* Evaluation of a portable device to measure daily energy expenditure. *Am Soc Nutr* 2007;**85**:742–9. doi:10.1093/ajcn/85.3.742

81 Soric M, Mikulic P, Misigoj-Durakovic M, *et al.* Validation of the Sensewear Armband during recreational in-line skating. *Eur J Appl Physiol* 2012;**112**:1183–8. doi:10.1007/s00421-011-2045-6

82 Tucker WJ, Bhammar DM, Sawyer BJ, *et al.* Validity and reliability of Nike + Fuelband for estimating physical activity energy expenditure. *BMC Sports Sci Med Rehabil* 2015;**7**:14. doi:10.1186/s13102-015-0008-7

83 Vanhelst J, Mikulovic J, Bui-Xuan G, *et al.* Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. *BMC Res Notes* 2012;**5**:187. doi:10.1186/1756-0500-5-187

84 Van Hoye K, Mortelmans P, Lefevre J. Validation of the SenseWear Pro3 armband using an incremental exercise test. *J Strength Cond Res* 2014;**28**:2806–14. doi:10.1519/JSC.0b013e3182a1f836

85 Van Hoye K, Boen F, Lefevre J. Validation of the SenseWear Armband in different ambient temperatures. *J Sports Sci* 2015;**33**:1007–18. doi:10.1080/02640414.2014.981846

86 Vernillo G, Savoldelli A, Pellegrini B, *et al.* Validity of the SenseWear Armband to Assess Energy Expenditure in Graded Walking. *J Phys Act Heal* 2015;**12**:178–83. doi:10.1123/jpah.2013-0437

87 Wahl Y, Düking P, Droszez A, *et al.* Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Front Physiol* 2017;**8**:725. doi:10.3389/fphys.2017.00725

88 Wallen MP, Gomersall SR, Keating SE, *et al.* Accuracy of heart rate watches: Implications for weight management. *PLoS One* 2016;**11**:e0154420. doi:10.1371/journal.pone.0154420

89 Sanghvi A, Redman LM, Martin CK, *et al.* Validation of an inexpensive and accurate mathematical method to measure long-term changes in free-living energy intake. *Am J Clin Nutr* 2015;**102**:353–8. doi:10.3945/ajcn.115.111070

90 Pollard TM, Wagnild JM. Gender differences in walking (for leisure, transport and in total) across adult life: a systematic review. *BMC Public Health* 2017;**17**:341. doi:10.1186/s12889-017-4253-4

91 Schneller MB, Pedersen MT, Gupta N, *et al.* Validation of five minimally obstructive methods to estimate physical activity energy expenditure in young adults in semi-standardized settings. *Sensors (Basel)* 2015;**15**:6133–51. doi:10.3390/s150306133

92 Welk GJ, McClain JJ, Eisenmann JC, *et al.* Field Validation of the MTI Actigraph and BodyMedia Armband Monitor Using the IDEEA Monitor. *Obes* 2007;**15**:918–28.http://search.ebscohost.com/login.aspx?direct=true&db=sph&AN=24897397&site=ehost-live

93 A.H.K. M, J.M. P, L.M. M, *et al.* Comparison of Activity Type Classification Accuracy from Accelerometers Worn on the Hip, Wrists, and Thigh in Young, Apparently Healthy Adults. *Meas Phys Educ Exerc Sci* 2016;**20**:173–83. doi:http://dx.doi.org/10.1080/1091367X.2016.1192038

94 Plasqui G. Smart approaches for assessing free-living energy expenditure following identification of types of physical activity. *Obes Rev* 2017;**18**:50–5. doi:10.1111/obr.12506

95 Brage S, Ekelund U, Brage N, *et al.* Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol* 2007;**103**:682–92. doi:10.1152/japplphysiol.00092.2006

96 Benedetto S, Caldato C, Bazzan E, *et al.* Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One* 2018;**13**:e0192691. doi:10.1371/journal.pone.0192691

97 Higgins JPT. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* 2008;**37**:1158–60. doi:10.1093/ije/dyn204

98 Kipp S, Byrnes WC, Kram R. Calculating metabolic energy expenditure across a wide range of exercise intensities: the equation matters. *Appl Physiol Nutr Metab* 2018;:1–4. doi:10.1139/apnm-2017-0781

Legends:

Table 1. Moderation analysis for level of sensors and grade of device by subgroup. Data are shown where at least 3 comparisons were included. *P*-value refers to a between subgroup comparison. *Significant effect size at the subgroup level (p<.05). Abbreviations: Accelerometry alone (ACC), accelerometry and heart rate (ACC+HR), accelerometry and heart rate and heat sensing (ACC+HR+HS) and accelerometry and heat sensing (ACC+HS). Activity energy expenditure (AEE), Total energy expenditure (TEE), Doubly labelled water (DLW).

PLEASE INSERT FIGURE 1 AROUND LINE 216

Figure 1. Flow diagram of study selection.

PLEASE INSERT FIGURE 2 AROUND LINE 254

Figure 2. Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device over all activities. Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.

Abbreviations: *Actical (ACT), Actigraph GT3X (AGT3X), Apple watch (AW), Apple Watch series 2 (AWS2), Beurer AS80 (BA), Bodymedia CORE armband (BMC), Basis Peak (BP), Epson Pulsense (EP), ePulse Personal Fitness Assistant (EPUL), Fitbit Blaze (FB), Fitbit Charge (FC), Fitbit Charge 2 (FC2), Fitbit Charge HR (FCHR), Fitbit Flex (FF), Garmin Forerunner 225 (GF225), Garmin Forerunner 920XT (GF920XT), Garmin Vivoactive (GVA), Garmin vivofit (GVF), Garmin vivosmart (GVS), Garmin Vivosmart HR (GVHR), Jawbone UP (JU), Jawbone UP24 (JU24), LifeChek calorie sensor (LC), Mio Alpha (MA), Microsoft band (MB), Misfit Shine (MS), Nike Fuel band (NF), Polar Loop (PL), Polar: AW200 (PO200), Polar: AW360 (PA360), Samsung Gear S (SG), SenseWear Armband (SWA), SenseWear Armband Pro 2 (SWA p2), SenseWear Armband Pro 3 (SWA p3), SenseWear Armband MINI (SWAM), TOMTOM Touch (TT), Vivago (V), Withings Pulse (WP), Withings Pulse O2 (WPO).*

PLEASE INSERT FIGURE 3 AROUND LINE 284

Figure 3. Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device for ambulation and stair climbing. Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.

Abbreviations: *Actigraph GT3X (AGT3X), Apple watch (AW), Beurer AS80 (BA), Bodymedia CORE armband (BMC), Basis Peak (BP), ePulse Personal Fitness Assistant (EPUL), Fitbit Charge (FC), Fitbit Charge HR (FCHR), Fitbit Flex (FF), Garmin Forerunner 225 (GF225), Garmin Forerunner 920XT (GF920XT), Garmin Vivoactive (GVA), Garmin vivofit (GVF), Garmin vivosmart (GVS), Jawbone UP (JU), Jawbone UP24 (JU24), Microsoft band (MB), Nike Fuel band (NF), Polar Loop (PL), Polar: AW200 (PO200), SenseWear Armband (SWA), SenseWear Armband Pro 2 (SWA p2), SenseWear Armband Pro 3 (SWA p3), SenseWear Armband MINI (SWAM), Vivago (V), Withings Pulse (WP), Withings Pulse O2 (WPO).*

PLEASE INSERT FIGURE 4 AROUND LINE 313

Figure 4. Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device for sedentary and household tasks.

Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.
Abbreviations: *Apple watch (AW), Bodymedia CORE armband (BMC), Basis Peak (BP), ePulse Personal Fitness Assistant (EPUL), Fitbit Charge HR (FCHR), Fitbit Flex (FF), Garmin Forerunner 225 (GF225), Garmin vivofit (GVF), Jawbone UP (JU), Jawbone UP24 (JU24), Microsoft band (MB), SenseWear Armband Pro 2 (SWA p2), SenseWear Armband Pro 3 (SWA p3), SenseWear Armband MINI (SWAM), Vivago (V), Withings Pulse (WP).*

PLEASE INSERT FIGURE 5 AROUND LINE 320

Figure 5. Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device for total energy expenditure (TEE). Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.
Abbreviations: *Epson Pulsense (EP), Fitbit Flex (FF), Garmin vivofit (GVF), Jawbone UP24 (JU24), Misfit Shine (MS), SenseWear Armband (SWA), SenseWear Armband Pro 2 (SWA p2), SenseWear Armband Pro 3 (SWA p3), SenseWear Armband MINI (SWAM), Withings Pulse O2 (WPO).*

Figure 1:

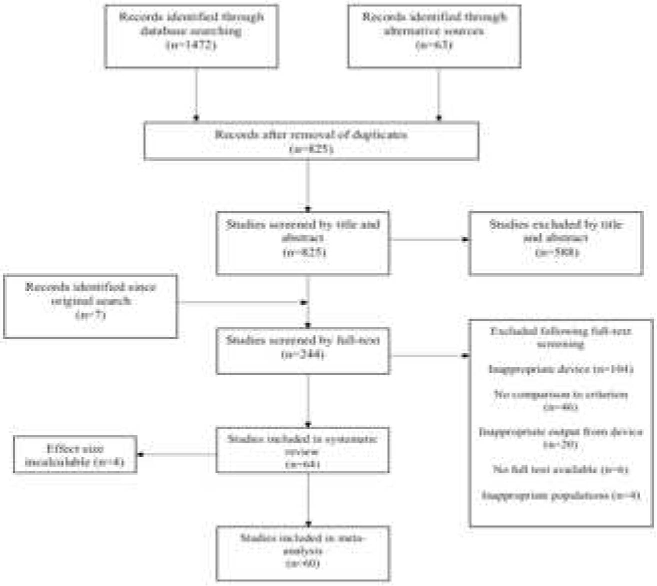
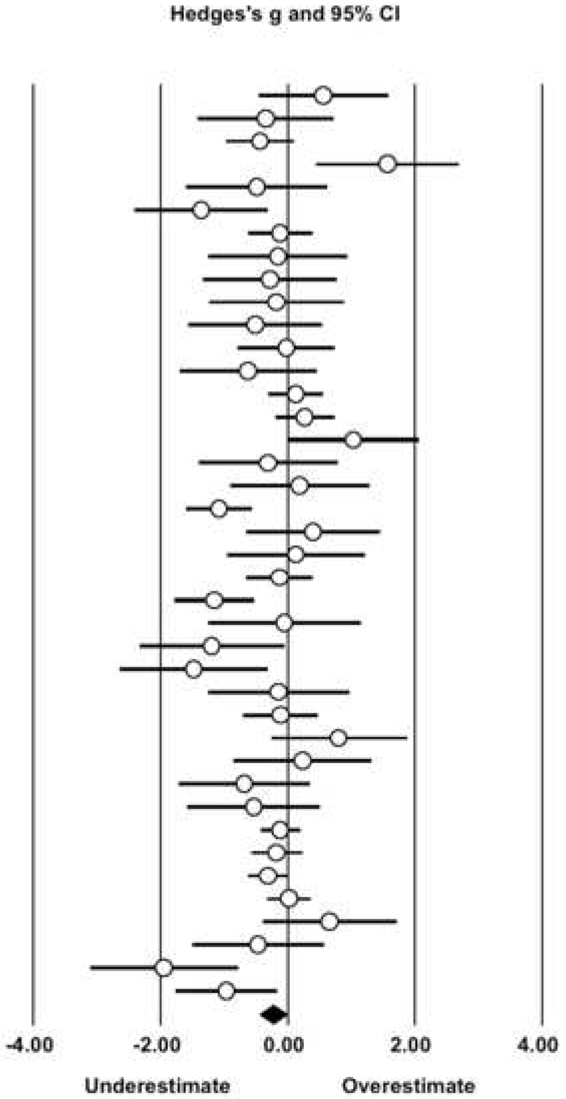

Records identified through database searching (n=1472)
Records identified through alternative sources (n=63)
Records after removal of duplicates (n=825)
Studies screened by title and abstract (n=825)
Studies excluded by title and abstract (n=588)
Records identified since original search (n=7)
Studies screened by full-text (n=244)
Excluded following full-text screening
Inappropriate device (n=104)
No comparison to criterion (n=46)
Inappropriate output from device (n=20)
No full text available (n=6)
Inappropriate populations (n=4)
Effect size incalculable (n=4)
Studies included in systematic review (n=64)
Studies included in meta-analysis (n=60)

Figure 2:

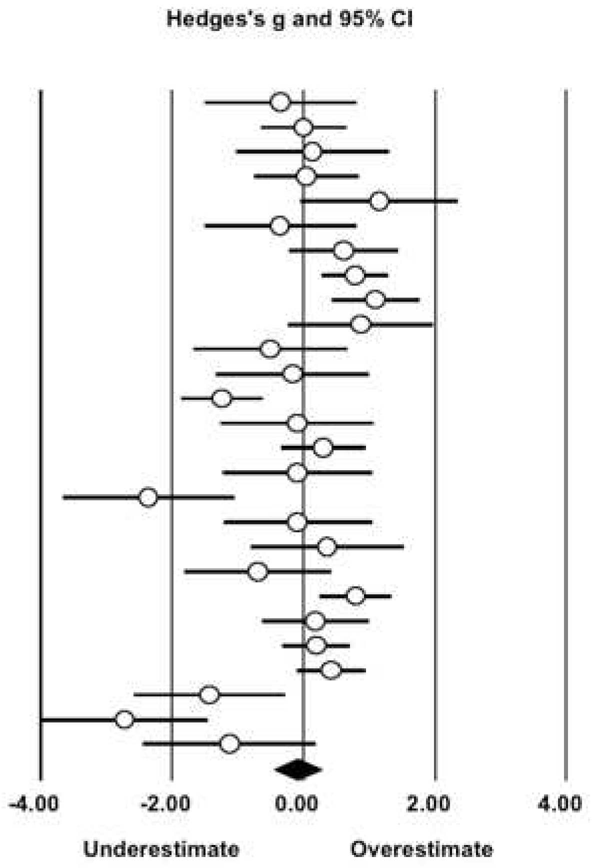
	Total	ES	Lower limit	Upper limit	p-Value
ACT	1	0.56	-0.46	1.58	0.28
AGT3X	1	-0.35	-1.42	0.73	0.53
AW	4	-0.43	-0.97	0.10	0.11
AWS2	1	1.58	0.45	2.70	0.01
BA	1	-0.49	-1.61	0.64	0.40
BB1	1	-1.37	-2.42	-0.31	0.01
BMC	4	-0.12	-0.64	0.40	0.65
BP	1	-0.15	-1.25	0.94	0.78
EP	1	-0.28	-1.34	0.78	0.60
EPUL	1	-0.18	-1.24	0.88	0.74
FB	1	-0.51	-1.57	0.54	0.34
FC	2	-0.02	-0.79	0.75	0.95
FC2	1	-0.62	-1.70	0.45	0.26
FCHR	6	0.13	-0.31	0.56	0.57
FF	5	0.27	-0.20	0.74	0.26
GF225	1	1.04	0.00	2.08	0.05
GF920XT	1	-0.31	-1.41	0.79	0.58
GVA	1	0.19	-0.91	1.29	0.74
GVF	5	-1.09	-1.60	-0.57	0.00
GVHR	1	0.40	-0.66	1.46	0.46
GVS	1	0.13	-0.97	1.23	0.81
JU	4	-0.13	-0.66	0.39	0.62
JU24	3	-1.16	-1.78	-0.54	0.00
LC	1	-0.05	-1.26	1.15	0.93
MA	1	-1.19	-2.34	-0.05	0.04
MB	1	-1.48	-2.64	-0.31	0.01
MS	1	-0.14	-1.26	0.98	0.80
NF	3	-0.12	-0.72	0.48	0.69
PA360	1	0.80	-0.27	1.88	0.14
PL	1	0.23	-0.85	1.32	0.67
Polar AW200	1	-0.69	-1.73	0.35	0.19
SG	1	-0.54	-1.59	0.51	0.31
SWA	12	-0.12	-0.43	0.19	0.45
SWA p2	7	-0.17	-0.57	0.22	0.39
SWA p3	12	-0.32	-0.62	-0.01	0.04
SWAM	9	0.02	-0.33	0.37	0.90
TT	1	0.67	-0.40	1.73	0.22
V	1	-0.47	-1.51	0.57	0.38
WP	1	-1.95	-3.12	-0.78	0.00
WPO	2	-0.97	-1.77	-0.16	0.02
Overall	104	-0.23	-0.44	-0.03	0.03



Meta Analysis Overall

Figure 3:

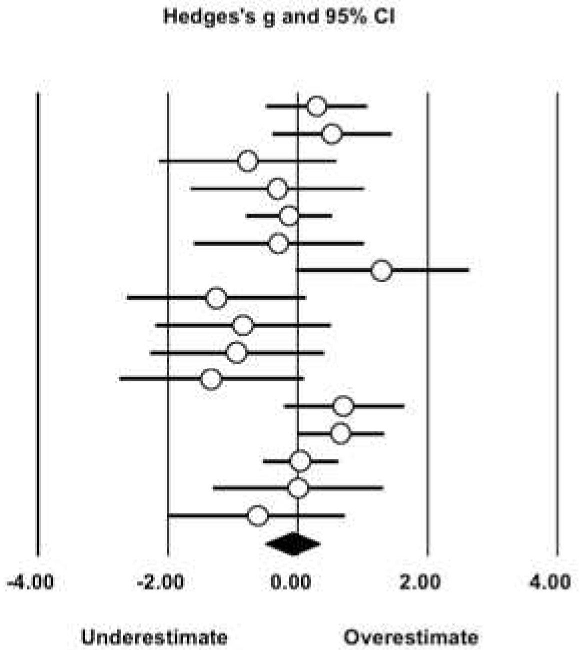
	Total	ES	Lower limit	Upper limit	p-Value
AGT3X	1	-0.35	-1.50	0.81	0.56
AW	3	0.00	-0.65	0.65	1.00
BA	1	0.14	-1.02	1.31	0.81
BMC	2	0.05	-0.75	0.85	0.90
BP	1	1.15	-0.04	2.35	0.06
EPUL	2	-0.35	-1.50	0.80	0.55
FC	2	0.61	-0.23	1.45	0.15
FCHR	5	0.78	0.28	1.29	0.00
FF	3	1.10	0.43	1.77	0.00
GF225	1	0.87	-0.24	1.98	0.12
GF920XT	1	-0.50	-1.68	0.67	0.40
GVA	1	-0.16	-1.33	1.00	0.78
GVF	4	-1.24	-1.86	-0.62	0.00
GVS	1	-0.10	-1.26	1.07	0.87
JU	3	0.31	-0.34	0.95	0.35
JU24	1	-0.09	-1.23	1.04	0.87
MB	1	-2.36	-3.67	-1.05	0.00
NF	1	-0.08	-1.21	1.04	0.88
PL	1	0.36	-0.81	1.53	0.54
Polar AW200	1	-0.69	-1.81	0.43	0.23
SWA	5	0.79	0.25	1.33	0.00
SWA p2	2	0.18	-0.63	0.99	0.67
SWA p3	5	0.20	-0.32	0.71	0.46
SWAM	5	0.43	-0.09	0.94	0.10
V	1	-1.42	-2.58	-0.27	0.02
WP	1	-2.72	-3.98	-1.46	0.00
WPO	1	-1.12	-2.44	0.20	0.10
Overall	55	-0.09	-0.45	0.27	0.62



Meta Analysis Ambulation and Stairs

Figure 4:

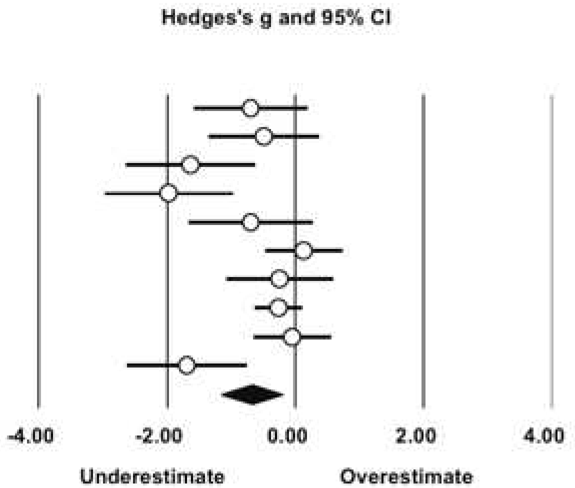
	Total	ES	Lower limit	Upper limit	p-Value
AW	3	0.29	-0.49	1.07	0.47
BMC	2	0.52	-0.41	1.45	0.27
BP	1	-0.78	-2.14	0.59	0.27
EPUL	1	-0.31	-1.65	1.02	0.65
FCHR	4	-0.14	-0.81	0.53	0.69
FF	1	-0.30	-1.62	1.02	0.66
GF225	1	1.30	-0.04	2.64	0.06
GVF	1	-1.26	-2.64	0.12	0.07
JU	1	-0.85	-2.20	0.51	0.22
JU24	1	-0.94	-2.28	0.40	0.17
MB	1	-1.34	-2.75	0.08	0.06
SWA p2	2	0.71	-0.23	1.64	0.14
SWA p3	4	0.67	0.00	1.34	0.05
SWAM	5	0.04	-0.55	0.63	0.90
V	1	0.00	-1.32	1.32	1.00
WP	1	-0.62	-1.97	0.73	0.37
Overall	30	-0.09	-0.51	0.32	0.66



Meta Analysis Sedentary and Household

Figure 5:

	Total	ES	Lower limit	Upper limit	p-Value
EP	1	-0.68	-1.58	0.21	0.16
FF	1	-0.48	-1.35	0.38	0.27
GVF	1	-1.62	-2.63	-0.62	0.00
JU24	1	-1.96	-2.97	-0.95	0.00
MS	1	-0.68	-1.65	0.29	0.17
SWA	2	0.14	-0.46	0.75	0.65
SWA p2	1	-0.23	-1.07	0.60	0.59
SWA p3	5	-0.25	-0.64	0.13	0.19
SWAM	2	-0.04	-0.64	0.56	0.90
WPO	1	-1.68	-2.62	-0.75	0.00
Overall	16	-0.68	-1.15	-0.21	0.00



Meta Analysis TEE (DLW)

S1:

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5 & 6
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	5 & 6
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	5 & 6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	S2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	7
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	7-8
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	8
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	

S2:

Population: Healthy adult populations (>18). Free from factors that impact physical movement.
Intervention: activity monitors + all research grade accelerometers (must be wearable on wrist or arm)
Comparison: Validated method: metabolic cart, DLW, DC, all IC systems,
Outcome: validity of energy expenditure (kcal/kj/met/correlation),

	P	I	C	O
Key concepts	**ADULTS**	**ACTIVITY MONITORS**	**VALIDATED METHOD**	**ENERGY EXPENDITURE**
Related terms		FITNESS TRACKERS (CINHAL) ACCELEROMETRY (MESH) ACCELEROMETER AMBULATORY MONITOR* FITBIT ACTIVITY MONITOR	VALID* COMPAR* TEST	ENERGY METABOLISM (MESH) CALORIES ENERGY EXPENDITURE CALORIC EXPENDITURE TOTAL DAILY ENERGY EXPENDITURE TDEE AEE
Terms to include in search		1. Activity tracker 2. Activity Monitor 3. Health tracker 4. Health monitor 5. Fitness tracker 6. Fitness monitor 7. Physical activity tracker 8. Physical activity monitor 9. Exercise tracker 10. Exercise monitor	1. Doubly labelled water 2. Dlw 3. Indirect caliomet* 4. Caliomet* 5. Direct caliomet* 6. Metabolic chamber 7. Metabolic cart 8. Gold standard 9. Criterion	Energy expenditure 1. Energy metabolism 2. Calori* 3. Calori* expenditure 4. Total energy expenditure 5. Activity energy expenditure 6. AEE 7. TDEE

		11. Electronic tracker 12. Electronic monitor 13. acceleromet 14. Step tracker 15. Wearable		

(Tracker AND EE) AND Validation

1. Activity tracker
2. Activity Monitor
3. Health tracker
4. Health monitor
5. Fitness tracker
6. Fitness monitor
7. Physical activity tracker
8. Physical activity monitor
9. Exercise tracker
10. Exercise monitor
11. Electronic tracker
12. Electronic monitor
13. acceleromet
14. Step tracker
15. Wearable

AND

1. Energy expenditure
2. Energy metabolism
3. Calori*
4. Calori* expenditure
5. Total energy expenditure
6. Activity energy expenditure
7. AEE
8. TDEE

AND

1. Doubly labelled water
2. Dlw
3. Indirect caliomet*
4. Caliomet*
5. Direct caliomet*
6. Metabolic chamber
7. Metabolic cart
8. Gold standard
9. Criterion

Database	Search	Results
Sport discus	**(activity tracker or activity monitor or health tracker or health monitor or fitness tracker or fitness monitor or physical activity tracker or physical activity monitor or exercise tracker or exercise monitor or electronic tracker or electronic monitor or acceleromet* or step tracker or wearable tracker) AND (energy expenditure or energy metabolism or calori* or calori* expenditure or total energy expenditure or activ* energy expenditure or AEE or TDEE) AND (doubly labelled water or DLW or indirect caliomet* or caliomet* or direct caliomet* or metabolic chamber or metabolic cart or gold standard or criterion)**	**154**
Pubmed	*((((((((((((((((((activity tracker) OR activity monitor) OR health tracker) OR health monitor) OR fitness trackers) OR fitness monitor) OR physical activity tracker) OR physical activity monitor) OR exercise trained) OR exercise monitor) OR electronic trackers) OR electronic monitor) OR acceleromet*) OR step tracer) OR wearable trackers)) AND ((((((((energy expenditure) OR energy metabolism) OR calori*) OR calori* expenditure) OR total energy expenditure) OR activ* energy expenditure) OR AEE) OR tdee))) AND (((((((((doubly labelled water) OR DLW) OR indirect caliomet*) OR caliomet*) OR direct caliomet*) OR metabolic chamber) OR metabolic cart) OR gold standard) OR criterion).*	**605**
MEDLINE	((activity tracker or activity monitor or health tracker or health monitor or fitness tracker or fitness monitor or physical activity tracker or physical activity monitor or exercise tracker or exercise monitor or electronic tracker or electronic monitor or acceleromet* or step tracker or wearable tracker).mp.	**228**

	AND (energy expenditure or energy metabolism or calori* or calori* expenditure or total energy expenditure or activ* energy expenditure or AEE or TDEE).mp. **AND** (doubly labelled water or DLW or indirect caliomet* or caliomet* or direct caliomet* or metabolic chamber or metabolic cart or gold standard or criterion).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word]	
Psycinfo	((activity tracker or activity monitor or health tracker or health monitor or fitness tracker or fitness monitor or physical activity tracker or physical activity monitor or exercise tracker or exercise monitor or electronic tracker or electronic monitor or acceleromet* or step tracker or wearable tracker).mp. **AND** (energy expenditure or energy metabolism or calori* or calori* expenditure or total energy expenditure or activ* energy expenditure or AEE or TDEE).mp. **AND** (doubly labelled water or DLW or indirect caliomet* or caliomet* or direct caliomet* or metabolic chamber or metabolic cart or gold standard or criterion).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word]	**26**
Embase	((activity tracker or activity monitor or health tracker or health monitor or fitness tracker or fitness monitor or physical activity tracker or physical activity monitor or exercise tracker or exercise monitor or electronic tracker or electronic monitor or acceleromet* or step tracker or wearable tracker).mp. **AND** (energy expenditure or energy metabolism or calori* or calori* expenditure or total energy expenditure or activ* energy expenditure or AEE or TDEE).mp.	**317**

	AND (doubly labelled water or DLW or indirect caliomet* or caliomet* or direct caliomet* or metabolic chamber or metabolic cart or gold standard or criterion).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word]	
CINHAL	**(activity tracker or activity monitor or health tracker or health monitor or fitness tracker or fitness monitor or physical activity tracker or physical activity monitor or exercise tracker or exercise monitor or electronic tracker or electronic monitor or acceleromet* or step tracker or wearable tracker) AND (energy expenditure or energy metabolism or calori* or calori* expenditure or total energy expenditure or activ* energy expenditure or AEE or TDEE) AND (doubly labelled water or DLW or indirect caliomet* or caliomet* or direct caliomet* or metabolic chamber or metabolic cart or gold standard or criterion)**	**142**
Obtained from reference lists		**63**
		AFTER REMOVAL OF DUPLICATES: 825

Exclusions:

1 = not comparison to criterion
2 = not comparison to accelerometer
3 = not healthy adult population
4 = review
5 = not kcal/kj
6= duplicate

S3:

	aee	walk	bike	run	sedentary and household	tee
Actical	0.77	0.8	0.3	0.8	0.7	0.7
Actigraph GT3X+	0.72	0.72	0.68	0.72	0.88	0.88
Apple watch	0.79	0.54	0.53	0.65	0.46	0.46
Basis b1	0.51	0.55	0.45	0.55	0.51	0.51
Basis Peak	0.51	0.55	0.45	0.66	0.49	0.49
Beurer AS80	0.44	0.44	0.44	0.44	0.44	0.44
BodyMedia FIT CORE	0.73	0.72	0.66	0.73	0.77	0.77
Epson Pulsense	0.71	0.71	0.71	0.71	0.71	0.71
ePulse Personal Fitness Assistant (ePulse)	0.24	0.24	0.24	0.24	0.24	0.24
Fitbit charge	0.32	0.68	0.44	0.68	0.41	0.41
Fitbit charge HR	0.77	0.75	0.53	0.68	0.41	0.41
Fitbit Flex	0.8	0.8	0.71	0.8	0.71	0.71
Fitbit Surge	0.77	0.75	0.53	0.68	0.41	0.41
Garmin Forerunner 225	0.35	0.35	0.35	0.35	0.35	0.35
Garmin Forerunner 920XT	0.35	0.35	0.35	0.35	0.35	0.35
Garmin vivoactive	0.75	0.75	0.75	0.35	0.75	0.75
Garmin vivofit	0.75	0.75	0.75	0.35	0.75	0.75
Garmin Vivosmart	0.75	0.75	0.75	0.35	0.75	0.75
Jawbone UP	0.82	0.8	0.73	0.74	0.53	0.53
Jawbone UP24	0.69	0.69	0.69	0.69	0.69	0.69
LifeChek calorie sensor	0.45	0.45	0.45	0.45	0.45	0.45
Microsoft band	0.54	0.55	0.46	0.54	0.44	0.44
Mio Alpha	0.46	0.46	0.46	0.46	0.46	0.46
Misfit Shine	0.41	0.41	0.41	0.41	0.41	0.41
Nike Fuel Band	0.77	0.77	0.77	0.77	0.77	0.77
Polar Loop	0.46	0.46	0.47	0.46	0.46	0.46
Polar: Activity Watch 200 (AW200)	0.7	0.7	0.7	0.7	0.7	0.7
PulseOn	0.45	0.45	0.4	0.45	0.45	0.45
Samsung Gear S	0.76	0.76	0.76	0.76	0.76	0.76
SenseWear Armband	0.73	0.72	0.66	0.73	0.77	0.77
SenseWear Mini Armband	0.73	0.72	0.66	0.73	0.77	0.77
SenseWear Pro 2 Armband	0.73	0.72	0.66	0.73	0.77	0.77
SenseWear Pro 3 Armband	0.73	0.72	0.66	0.73	0.77	0.77
TOM TOM TOUCH	0.2	0.3	0.3	0.3	0.3	0.3
Vivago	0.79	0.79	0.79	0.79	0.79	0.79
Withings Pulse	0.71	0.71	0.71	0.71	0.71	0.71
Withings Pulse 02	0.78	0.78	0.78	0.78	0.78	0.78

Correlations imputed for specific devices and activities

S4:

Study	Sample characteristics	Study protocol	Setting (Lab/ Field)	Criterion comparison	Device	Device placement	Results (overall error relative to criterion)
Alsubheen, 2016	N=13 (5 F) Age: 40 ± 11.9 y	Subjects performed a graded	Lab	IC – Sable system (Sable Systems	Garmin vivofit (Garmin ltd, Olathe,	Wrist	Garmin vivofit: -41.63%

	BMI: 27 ± 4.3 kg/m^2	treadmill test.		International, Las Vegas NV)	Kansas, USA)		
Bai, 2017	N=39 (16 F) Age: 32 ± 11 y BMI: 24.7 ± 4 kg/m^2	Subjects performed a semi-structured activity protocol consisting of sedentary activity, aerobic exercise, and light intensity physical activity on a treadmill.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	Apple watch (Apple Inc, Cupertino, California, USA) Fitbit charge HR (Fitbit Inc, San Francisco, California, USA)	Wrist	Apple Watch: -10.79% Fitbit Charge HR: 17.88%
Benito, 2012	N=29 (17 F) Age: 22.5 y BMI: 22 kg/m^2	Subjects performed circuits of resistance exercise at 30%, 50% and 70% of 15 repetition maximum.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Pro2 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro2 Armband: -46.60%
Berntsen, 2010	N=20 (6 F) Age: 35 y BMI: 24 kg/m^2	Subjects performed lifestyle and sporting activities including strength exercises, ball games, occupational and home-based activities.	Lab	IC – MetaMax II (Cortex Biophysic, Leipzig, Germany)	SenseWear Pro2 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro2 Armband: -9.00%
Berntsen, 2012	N=29 (29 F) Age: 31 ± 4.1 y BMI: 27 ± 3.2 kg/m^2	Subjects participated in a period of sedentary behaviour. 9 subjects then performed callisthenics and cycling on a bicycle ergometer. The other 20 subjects performed outdoor walking followed by	Lab	IC – MetaMax II (Cortex Biophysic, Leipzig, Germany)	SenseWear Pro2 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro2 Armband: -10.34%

		relaxing, cycling and callisthenics.					
Bhammar, 2016	N=34 (26 F) Age: 30.1 ± 8.7 y BMI: 26.2 ± 5.1 kg/m^2	Subjects performed a semi structured and a structured routine. Semi-structured: 12 activities including 4 sedentary/light-intensity activities, 4 moderate-intensity activities, and 4 vigorous-intensity activities. The activities performed were randomly selected from a list of common activities. Structured: A period of rest, followed by 7 activities of 8 minutes each. The activities performed were randomly selected from a list of common activities.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Mini Armband: 14.76%
Boudreaux, 2018	N=50 (28 F) Age: 22.4 y BMI: 26.5 kg/m^2	Subjects performed separate trials of graded cycling and 3 sets of 4 resistance exercises at a 10-	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	Apple Watch 2 (Apple Inc, Cupertino, California, USA) Fitbit Blaze (Fitbit Inc, San		Apple Watch 2: 48.20% Fitbit Blaze: 28.66% Fitbit Charge 2: -30.97%

		repetition maximum load.			Francisco, California, USA) Fitbit Charge 2 (Fitbit Inc, San Francisco, California, USA) Garmin Vivosmart HR (Garmin ltd, Olathe, Kansas, USA) Polar: the Activity Watch 360 (Polar Electro Oy, Kempele, Finland) Tomtom touch (TomTom, Amsterdam, the Netherlands)		Garmin Vivosmart HR: 16.85% Polar: the Activity Watch 360: 28.68% Tomtom Touch: 28.66%
Brazeau, 2011	N=31 (16 F) Age: 26.7 y BMI: 27.5 kg/m^2	Subjects performed 45 minutes of stationary cycling at 50% VO_{2peak}.	Lab	IC – Ergocard exercise test station (MediSoft, Dinant, Belgium)	SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro3 Armband: -10.56%
Brazeau, 2014	N=38 (18 F) Age: 28.6 y BMI: 23.8 kg/m^2	Subjects performed 45 minutes of treadmill exercise at 40% VO_{2peak} then exercised on a stationary bike ergometer for 45 minutes at 50% VO_{2peak}.	Lab	IC – Ergocard exercise test station (MediSoft, Dinant, Belgium)	SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro3 Armband: 14.94%

Brazeau, 2016	N=20 (0 F) Age: 26.2 ± 3.6 y BMI: 23.1 ± 2.3 kg/m^2	Subjects completed a field observation and a lab protocol. Field: 7-day comparison to DLW. Lab: Subjects performed 60 minutes rest followed by treadmill exercise for 45 minutes at 22-41% VO_{2peak} then stationary cycling for 45 minutes at 50% VO_{2peak}.	Lab/ Field	DLW – 7 days IC – Ergocard exercise test station (MediSoft, Dinant, Belgium)	SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro3 Armband: 7.06%
Brugniaux, 2010	N=31 (16 F) Age: 42.9 y BMI: 22.7 kg/m^2	Subjects performed a 9.7km outdoor hike.	Field	IC – Metablograph with Hans Rudolph facemask (Hans Rudolph, Kansas City, MO, USA)	Polar: the Activity Watch 200 (Polar Electro Oy, Kempele, Finland)	Wrist	Polar: the Activity Watch 200: -13.17%
Calabro, 2014	N=40 (19 F) Age: 27.4 y BMI: 22.8 kg/m^2	Subjects performed 60 minutes of structured activities including stationary biking, walking/ running on a treadmill, road biking, elliptical exercise and stair stepping and unstructured movements. The semi-structured measurement periods were	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Mini Armband: 0.89% SenseWear Pro3 Armband: 2.33%

		performed in 5, 10, 10, 10, and 25-minute intervals and included sitting, walking, standing, stair climbing or light movements.					
Calabro, 2015	N=29 (17 F) Age: 68.8 ± 6.3 y BMI: 26.3 ± 4.9 kg/m^2	14-day comparison to DLW.	Field	DLW – 14 days	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Mini Armband: -0.86%
Casiraghi, 2013	N=18 (11 F) Age: 48.6 ± 21 y BMI: 24.6 ± 2.6 kg/m^2	Subjects performed a cycling protocol with three components: 1) Baseline where the subject sat on the cycle ergometer. 2) A 2-minute warm-up at 40 rpm at 40 watts. 3) Exercise increased to 60 rpm and intensity progressed by 7 watts/minute until exhaustion.	Lab	IC – SensorMedics Vmax 229 (SensorMedics Inc, Yorba Linda, CA, USA).	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -8.00%
Chowdhry, 2017	N=30 (15 F) Age: 27 ± 1.6 y BMI: 23.4 ± 2.5 kg/m^2	Subjects performed two components: 1) A protocol of 4 activities of designed to replicate daily living tasks 2) 4 activities of	Lab	IC – COSMED K4b2 (COSMED, Rome, Italy)	Apple watch (Apple Inc, Cupertino, California, USA) Microsoft Band (Microsoft Corporation, Redmond,	Wrist Bodymedia core: Upper arm	Apple watch: -6.9% Microsoft Band: -49.15% Fitbit Charge HR: 15.49% Jawbone UP24:

		10 minutes in duration. These activities were walking on a treadmill, walking at the same speed with shopping bags, cycling on an ergometer and jogging on the treadmill.			Washington, USA) Fitbit Charge HR (Fitbit Inc, San Francisco, California, USA) Jawbone UP24 (Jawbone, San Francisco, California, USA) Bodymedia Core (HealthWear, Bodymedia, Pittsburg, PA, USA)		-21.01% Bodymedia Core: 7.98%
Colbert, 2011	N=56 (45 F) Age: 74.7 ± 6.5 y BMI: 25.8 ± 4.2 kg/m^2	10-day comparison to DLW.	Field	DLW – 10 days	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband: 58.53%
Correa, 2016	N=87 (72 F) Age: 42 ± 13 y BMI: 31.6 ± 4.5 kg/m^2	7-day comparison to DLW.	Field	DLW – 7 days	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) Actical (Phillips Respironics Inc, Murrysville, PN, USA)	Upper arm Wrist	SenseWear Armband −416.95 kcal Actical: 194.52 kcal
Diaz, 2015	N=23 (13 F) Age: N/A BMI: N/A	Subjects performed a treadmill protocol consisting of walking at slow, moderate and brisk	Lab	IC – Ultima CPX (Medgraphics, Saint Paul, MN, USA)	Fitbit Flex (Fitbit Inc, San Francisco, CA, USA)	Wrist	Fitbit Flex: 17.36%

		paces and jogging.					
Diaz, 2016	N=13 (13 F) Age: 32.0 ± 9.2 y BMI: 24.2 ± 3.4 kg/m^2	Subjects performed a treadmill protocol consisting of walking at slow, moderate and brisk paces and jogging.	Lab	IC – Ultima CPX (Medgraphics, Saint Paul, MN, USA)	Fitbit Flex (Fitbit Inc, San Francisco, CA, USA)	Wrist	Fitbit Flex: 30.27%
Dondzila, 2016	N=19 (5 F) Age: 24.6 ± 3.1 y BMI: 28.0 ± 3.8 kg/m^2	Subjects performed 5-minute stages of jogging on a treadmill at increasing velocity.	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	Fitbit Charge (Fitbit Inc, San Francisco, California, USA)	Wrist	Fitbit Charge: -13.01%
Dooley, 2017	N=62 (36 F) Age: 22.46 y BMI: 24.86 kg/m^2	Subjects performed 4 stages of treadmill exercise followed by a seated recovery period. The activity routine consisted of an unmeasured warm-up walking period and measured stages of slow, then brisk walking and jogging.	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	Apple watch (Apple Inc, Cupertino, CA, USA) Fitbit charge HR (Fitbit Inc, San Francisco, CA, USA) Garmin Forerunner 225 (Garmin ltd, Olathe, Kansas, USA)	Wrist	Apple watch: 64.55% Fitbit charge HR: 18.70% Garmin Forerunner 225: 44.23%
Drenowatz, 2011	N=20 (10 F) Age: 24.3 y BMI: N/A	Subjects performed three treadmill runs at 65, 75, and 85% VO_{2max}.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -32.80%
Erdogan, 2010	N=43 (27 F) Age: 34.9 ± 5.5 y	Subjects performed rowing exercises at 50% and	Lab	IC – COSMED K4b2 (COSMED, Rome, Italy)	SenseWear Armband (HealthWear, Bodymedia	Upper arm	SenseWear Armband: 5.23%

	BMI: 31.2 ± 3.7 kg/m^2	70% VO_{2max} on an ergometer.			, Pittsburgh, PA, USA)		
Fruin, 2010	Experiment 1: N=13 (0 F) Experiment 2: N=20 (10 F) Age: 20.2 ± 1 y BMI: N/A	Experiment 1: Subjects performed two resting and a cycle ergometer session at 60% VO_{2peak}. Experiment 2: Subjects completed a treadmill protocol of jogging, running and uphill running.	Lab	IC – SensorMedics Vmax 229 (SensorMedics Inc, Yorba Linda, CA, USA).	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -1.76%
Furlanetto, 2010	N=30 (15 F) Age: 68 ± 7 y BMI: 25 ± 3 kg/m^2	Subjects performed a walking protocol on a treadmill at three intensities.	Lab	IC – VO_{2000} aerograph (Medgraphics, Saint Paul, MN, USA)	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -6.99%
Gastin, 2017	N=26 (12 F) Age: 21.3 ± 2.4 y BMI: 23.2 ± 2 kg/m^2	Subjects performed a protocol Involving resting periods, walking, jogging, running or a sport-simulated circuit.	Lab	IC – MetaMax 3b (Cortex Biophysic, Leipzig, Germany)	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -19.90%
Heiermann, 2011	N=32 (19 F) Age: 68.6 y BMI: 26.4 kg/m^2	Subjects were required to rest.	Lab	IC – Vmax Spectra (SensorMedics Viasys Healthcare, Bilthoven, The Netherlands)	SenseWear Pro2 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro2 Armband: 10.80%
Imboden, 2017	N=30 (15 F) Age: 49.2 ± 19.2 y BMI: 26.2 kg/m^2	Subjects performed a semi-structured activity protocol, performing ≥12 activities for	Lab	IC – COSMED K4b2 (COSMED, Rome, Italy)	Fitbit flex (Fitbit Inc, San Francisco, California, USA) Jawbone UP24	Wrist	Fitbit flex: -15.29% Jawbone UP24: -40.00%

		subject-selected duration and pace. Activities were selected from a list of sedentary, household activities ambulatory and cycling activities.			(Jawbone, San Francisco, California, USA)		
Jakicic, 2004	N=40 (20 F) Age: 23.2 ± 3.8 y BMI: 23.8 ± 3.1 kg/m^2	Subjects performed 4 separate exercise protocols including treadmill walking, stair stepping, cycle ergometry, and arm ergometry.	Lab	IC – SensorMedics Vmax 229 (SensorMedics Inc, Yorba Linda, CA, USA).	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -11.76%
Johannsen, 2010	N=30 (15 F) Age: 38.2 ± 10.6 y BMI: 24 ± 3.4 kg/m^2	14-day comparison to DLW.	Field	DLW – 14 days	SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) SenseWear Mini Armband HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro3 Armband: -2.48%
Kim, 2015	N=52 (19 F) Age: 23.8 ± 5.2 BMI: N/A	Subjects performed 15 activities including resting, stair climbing, cycling, walking and jogging. Each activity was performed for 5	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	Bodymedia Core (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	Bodymedia Core: 5.80%

		minutes, with 1-minute resting intervals.					
King, 2004	N=21 (10 F) Age: 37.55 y	Subjects performed 10 minutes of treadmill walking and running at various speeds.	Lab	IC – TrueMax 2400 (Consentius Technologies, Sandy, UT, USA)	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: 20.33%
Koehler. 2011	N=14 (0 F) Age: 30.4 ± 6.2 y BMI: 23.2 ± 1.4 kg/m^2	7-day comparison to DLW.	Field	DLW – 7 days	SenseWear Pro3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro3 Armband: -1.83%
Lee, 2011	N=46 (21 F) Age: 24.8 ± 5.6 y BMI: 24.3 ± 3.6 kg/m^2	Subjects completed 4-minute periods of standing, walking, jogging, and running.	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	ePulse Personal Fitness Assistant (ePulse) (Impact Sports Technologies, San Diego, CA, USA)	Forearm	ePulse Personal Fitness Assistant -3.46%
Lee, 2014	N=60 (30 F) Age: 26.4 y BMI: 23.05 kg/m^2	Subjects performed 13 activities for 5 minutes. Activities were categorized into sedentary, treadmill walking, treadmill jogging and moderate-to-vigorous activities (ascending and descending stairs, stationary bike, elliptical exercise, Wii tennis play,	Lab	IC – Oxycon Mobile 5.0 (Erich Jaeger, Viasys Healthcare, Germany)	BodyMedia CORE (BodyMedia Inc., Pittsburgh, PA, USA) Jawbone UP (Jawbone, San Francisco, California, USA) Basis B1 Band (Basis Science Inc, San Francisco, CA, USA) Nike Fuel Band (Nike Inc.,	Upper arm Wrist	BodyMedia CORE:-5.31% Jawbone UP: -6.92% Basis B1 Band: -31.65% Nike Fuel Band: -1.91%

		and basketball).			Beaverton, OR, USA)		
Lopez, 2017[1]	N=36 (16 F) Age: 37.7 ± 9.8 y BMI: 23.4 ± 2.8 kg/m^2	Subjects performed a structured protocol including rest, computer use, standing, slow walking, running, basketball and overground cycling.	Lab	IC – MetaMax 3x (Cortex Biophysic, Leipzig, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Mini Armband: -16.00%
Mackey, 2011	N=19 (8 F) Age: 82 ± 3.3 y BMI: 28.1 ± 3.8 kg/m^2	12.5-day comparison to DLW.	Field	DLW – 12.5 days	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: -0.05%
Martien, 2015	N=60 (47 F) Age: 85.5 ± 5.5 y BMI: N/A	Subjects performed activity for 4 minutes and separated by 4 minutes seated rest. Activities included: Walking, rising and sitting in chairs positioned 5 meters apart and moving light objects.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Mini Armband: -12.00%

Maschac, 2013[1]	N=19 (13 F) Age: 55.65 y BMI: 31.5 ± 3.6 kg/m^2	Subjects performed three walking sessions on a treadmill with different combinations of speed and incline.	Lab	IC – VO_{2000} aerograph (Medgraphics, Saint Paul, MN, USA)	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband: 50.69%
McMinn, 2013	N=19 (6 F) Age: 30 y BMI: 23.6 kg/m^2	Subjects completed 3 treadmill walking trials at self-selected slow, medium, and fast speeds.	Lab	IC – Ultima CPX (Medgraphics, Saint Paul, MN, USA)	Actigraph GT3X+ (Actigraph Inc, Pensacola, FL, USA)	Wrist	Actigraph GT3X+ : -8.84%
Melanson, 2009	N=7 (3 F) Age: 31.8 ± 7.2 y BMI: 27.8 ± 7.9 kg/m^2	Subjects performed individualised protocols, including bench stepping and stationary cycling.	Lab	MC – 22.8 hours	LifeChek Calorie Sensor (LifeChek, LLC, Pittsburgh, PA, USA)	Wrist	LifeChek calorie sensor -4.87%
Mikulic, 2011	N=19 (11 F) Age: 28 ± 6 y BMI: 23 ± 3 kg/m^2	Subjects performed in-line skating exercises on a circular track at a self-selected pace.	Field	IC – COSMED K4b2 (COSMED, Rome, Italy)	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband : -73.33%
Montoye, 2017	N=32 (14 F) Age: 23.7 y BMI: 25.5 kg/m^2	Subjects completed 14 exercises, 11 in the laboratory including walking, jogging and cycling ergometry and 3 track exercises included self-paced walking at both a leisure and brisk pace for 200 meters and self-paced jogging for	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	Fitbit Charge HR (Fitbit Inc, San Francisco, California, USA)	Upper arm	Fitbit Charge HR: 7.59%

		400 meters. Each was 5 minutes in duration.					
Murakami, 2016	N=19 (10 F) Age: N/A BMI: N/A	1) 12.5-day comparison to DLW. 2) 24 hours in metabolic chamber where subjects where subjects were required to perform deskwork, watch television, housework, treadmill walking, and sleeping.	Lab/ Field	DLW – 12.5 days MC – 24 hours	Withings Pulse O2 (Withings, Issy-les-Moulineaux, France) Garmin vivofit (Garmin ltd, Olathe, Kansas, USA) Fitbit Flex (Fitbit Inc, San Francisco, California, USA) Misfit Shine (Misfit, San Francisco, California, USA) Epson Pulsense (Epson, Suwa, Nagano Prefecture, Japan)	Wrist	Withings Pulse O2: -22.03% Garmin vivofit: -20.55% Fitbit Flex: -1.04% Misfit Shine: -2.36% Epson Pulsense: -4.28%
Nelson, 2016	N=30 (15 F) Age: 48.9 ± 19.4 y BMI: 26.3 ± 5.2 kg/m^2	Subjects performed a structured protocol consisting of sedentary, household, and ambulatory activities.	Lab	IC – COSMED K4b2 (COSMED, Rome, Italy)	Jawbone UP (Jawbone, San Francisco, California, USA) Fitbit Flex (Fitbit Inc, San Francisco, California, USA)	Wrist	Jawbone UP: -2.12% Fitbit Flex: 12.74%
Papazoglou, 2006	N=29 Age: N/A BMI: N/A	Subjects performed a resting	Lab	IC – SensorMedics Vmax	SenseWear Pro 2 Armband	Wrist	SenseWear Pro 2

		protocol in a larger sample and 29 of the obese subjects participated in low intensity modes of exercise including cycle ergometry, stair stepping and treadmill walking.		229 (SensorMedics Inc, Yorba Linda, CA, USA)	(HealthWear, Bodymedia, Pittsburgh, PA, USA)		Armband: 21.54%
Price, 2017	N=14 (3 F) Age: 23 y BMI: 22.8 kg/m^2	Subjects walked on a treadmill at increasing velocities.	Lab	IC – Parvo TrueOne 2400 (Parvo Medics, East Sandy, UT, USA)	Jawbone UP (Jawbone, San Francisco, California, USA) Garmin vivofit (Garmin ltd, Olathe, Kansas, USA)	Upper arm	Jawbone UP: 56.91% Garmin vivofit: 18.16%
Reece, 2015	N=22 (11 F) Age: N/A BMI: N/A	Subjects performed a protocol including rest, sedentary activities and walking.	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Wrist	SenseWear Mini Armband: -3.79%
Reeve, 2014[1]	N: 18 (7 F) Age: 22.6 y BMI: 22.9 kg/m^2	Subjects performed 2 resistance training sessions that included 9 different exercises. The weight lifted was 70% of 1 repetition max with 90-second rest intervals.	Lab	IC – COSMED K4b2 (COSMED, Rome, Italy)	BodyMedia CORE (BodyMedia Inc., Pittsburgh, PA, USA) SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	BodyMedia CORE: 13.8% SenseWear Mini Armband: 23.7%

Rousset, 2015	Free-living: N=41 (20 F) Lab: N=49 (26 F) Age: N/A BMI: N/A	1) 10-day comparison to DLW. 2) 24 hours in metabolic chamber, which included eating, deskwork, watching television, housework, treadmill walking, and sleeping.	Lab/ Field	DLW – 12.5 days MC – 17 hours	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband: -2.80%
Shcherbina, 2017[1]	N=60 (31 F) Age: 38.5 y BMI: 23.65 kg/m^2	Subjects performed treadmill flat and incline running and cycle ergometry at low and moderate intensity.	Lab	IC – COSMED Quark CPNET (COSMED, Rome, Italy)	Apple watch (Apple Inc, Cupertino, CA, USA) Basis Peak (Basis Science Inc, San Francisco, CA, USA) Fitbit surge (Fitbit Inc, San Francisco, CA, USA) Microsoft band (Microsoft Corporation, Redmond, WA, USA) PulseOn (PulseOn Oy, Espoo Finland)	Wrist	Apple watch: -38.23% Basis Peak: -12.94% Fitbit Surge: -3.86% Microsoft Band -19.64% PulseOn: -24.47%
Slinde, 2013	N=62 (62 F) Age: 33.2 ± 4.2 y	7-day comparison to DLW	Field	DLW – 7 days	SenseWear Pro 2 Armband (HealthWear,	Wrist	SenseWear Pro 2 Armband: -2.90%

	BMI: 30 ± 2.8 kg/m^2				Bodymedia, Pittsburgh, PA, USA)		
Smith, 2012	N=30 (30 F) Age: 29.0 ± 4.3 y BMI: 24.1 ± 3.0 kg/m^2	Subjects performed a series of activities of daily living activities and treadmill walking at increasing intensities.	Lab	IC – Parvo TrueOne 2400 (Parvo Medics East Sandy, UT, USA)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) Algorithm v2.2	Upper arm	SenseWear Mini Armband: 18.43%
Stackpool, 2014	N=20 (10 F) Age: N/A BMI: N/A	Subjects performed treadmill walking, treadmill running, elliptical exercise and an agility drills.	Lab	IC – Oxycon pro Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	Nike Fuel Band (Nike Inc, Beaverton, OR, USA) Jawbone UP (Jawbone, San Francisco, California, USA) Bodymedia Core (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	Nike Fuel Band: -3.99% Jawbone UP: 3.09%
St-Onge, 2007	N=45 (32 F) Age: 35.1 ± 14 y BMI: 23.9 ± 4.0 kg/m^2	10-day comparison to DLW.	Field	DLW – 10 days	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Armband: 4.70%
Tucker, 2015	N=24 (13 F) Age: 28.4 ± 7.8 y BMI: 23.8 ± 3.9 kg/m^2	Subjects performed two, 60-minute semi-structured routines consisting of sedentary/light-intensity, moderate-intensity and	Lab	IC – Oxycon Mobile portable metabolic system (Erich Jaeger, Viasys Healthcare, Germany)	Nike Fuel Band (Nike Inc., Beaverton, OR, USA) SenseWear Armband (HealthWear, Bodymedia	Upper arm	Nike Fuel Band: 1.22% SenseWear Armband: -2.10%

		vigorous-intensity physical activity.			, Pittsburgh, PA, USA)		
Van Helst, 2012	N=21 (10 F) Age: 29.3 ± 5.1 y	Subjects performed a treadmill protocol involving slow and moderate walking, running slowly, vigorously running and periods of rest.	Lab	IC – Gas analyzer (Respironics Novametrix Medical SystemW inc, NICO 7300, Wallingford, USA)	Vivago (Vivago Wellness, Paris, France)	Wrist	Vivago: -8.02%
Van Hoye, 2014	N=44 (20 F) Age: 21.1 ± 1.4 y BMI: 21.8 ± 1.4 kg/m^2	Subjects performed an incremental running test on a treadmill.	Lab	IC – Metalyzer 3B (Cortex Biophysic, Leipzig, Germany)	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband: -32.96%
Van Hoye, 2015	N=39 (18 F) Age: 21.1 ± 1.4 y BMI: 21.8 ± 1.4 kg/m^2	Subjects performed exercise consisting of 5 minutes standing followed by alternating walking and running at 35% and 65% VO_{2max}.	Lab	IC – Metalyzer 3B (Cortex Biophysic, Leipzig, Germany)	SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) Algorithm v2.2 SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) Algorithm v5.2	Upper arm	SenseWear Pro 3 Armband: --15.23%
Vernillo, 2015	N=20 (8 F) Age: 30.1 ± 7.2 y BMI: 22.1 ± 2.4 kg/m^2	Subjects performed randomized pole walking activities at a constant speed and a variety of gradients.	Lab	IC – COSMED Quark b2 (COSMED, Rome, Italy)	SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	Upper arm	SenseWear Pro 3 Armband: -9.76% SenseWear Mini Armband: -12.50

					SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)		
Wahl, 2017	N=20 (10 F) Age: 25.2 y BMI: 22.8 kg/m^2	Subjects performed a running protocol consisting of four 5-minute stages of treadmill running at different velocities followed by a period of intermittent running and then a 2.4 km outdoor run.	Lab/ Field	IC – Metalyzer 3B (Cortex Biophysic, Leipzig, Germany)	SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA) Beurer AS80 (Beurer GmbH, Ulm, Germany) Polar Loop (Polar Electro, Kempele, Finnland) Garmin vivofit (Garmin ltd, Olathe, Kansas, USA) Garmin vivosmart (Garmin ltd, Olathe, Kansas, USA) Garmin vivoactive (Garmin ltd, Olathe, Kansas, USA) Garmin Forerunner 920XT (Garmin ltd, Olathe,	Upper arm/Wrist	SenseWear Mini Armband: -21.27% Beurer AS80: -58.07% Polar Loop: 18.05% Garmin vivofit: -13.67% Garmin vivosmart: 5.98% Garmin vivoactive: 3.42% Garmin Forerunner 920XT: -21.02% Fitbit Charge: 3.58% Fitbit charge HR: 7.58% Withings Pulse O2: -15.98%

					Kansas, USA) Fitbit Charge (Fitbit Inc, San Francisco, California, USA) Fitbit charge HR (Fitbit Inc, San Francisco, California, USA) Withings Pulse (Withings, Issy-les-Moulineaux, France)		
Wallen 2016	N=22 (11 F) Age: 24.9 y BMI: 24.3 kg/m^2	Subjects performed a protocol including treadmill exercise and cycling ergometry.	Lab	IC – Metalyzer 3B (Cortex Biophysic, Leipzig, Germany)	Apple watch (Apple Inc, Cupertino, California, USA) Fitbit charge HR (Fitbit Inc, San Francisco, California, USA) Samsung Gear S (Samsung Electronics Co, Ltd, Suwon, South Korea) Mio Alpha (Mio Global, Canada)	Wrist	Apple watch: -75.71 Fitbit charge HR: -26.31% Samsung Gear S: -9.98% Mio Alpha: -53.19%
Woodman, 2017	N=28 (8 F) Age: 24.85 y BMI: 24.25 kg/m^2	Subjects performed a range of activities including: supine rest, household	Lab/ Field	IC – Oxycon Mobile portable metabolic system (Erich	Withings Pulse (Withings, Issy-les-Moulineaux, France)	Wrist	Withings Pulse: -133.33% Basis Peak: 0.59%

tasks, treadmill walking, stair stepping, outdoor walking, cycling, and running at a self-selected pace. Seated rest, and ergometer cycling.	Jaeger, Viasys Healthcare, Germany)	Basis Peak (Basis Science Inc, San Francisco, CA, USA) Garmin vivofit (Garmin ltd, Olathe, Kansas, USA)	Garmin vivofit: -80.59%

Characteristics of studies meeting inclusion criteria of systematic review. Results represents the mean percentage error between device measurements and criterion measurements.
[1]Not included in meta-analysis.
Abbreviations: Female (F), body mass index (BMI), indirect calorimetry (IC), metabolic chamber (MC), doubly labelled water (DLW), Kilocalories (Kcal)

S5:

Device	Price	Wear site	Device grade	Input setup data	Sensors	Output	Battery life	Number of comparisons in meta-analysis	Weighted percent error
Actical (Phillips Respironics Inc, Murrysville, PN, USA)	€678 (incl. software)/ €321 (unit)	Hip, ankle, wrist	Research	Age, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Activity intensity Kcals, steps	194 days	1	
Actigraph GT3X+ (Actigraph Inc, Pensacola, FL, USA)	$250	Hip, ankle, wrist	Research	Age, Gender, Race, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Activity intensity Kcals, sleep, steps	31 days	1	-8.84%
Apple watch (Apple Inc, Cupertino, California, USA)	£249	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance tracking, Kcals, HR, minutes of brisk	18 Hours	4	-6.59%

						activity			
Apple watch 2 (Apple Inc, Cupertino, California, USA)	£315	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance tracking, Kcals, HR, minutes of brisk activity	18 Hours	1	48.20%
Basis b1 (Basis Science Inc, San Francisco, CA, USA)	£149	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors: Yes	Steps, distance, Kcals, HR, active minutes, sleep	5 days	1	-31.65%
Basis Peak (Basis Science Inc, San Francisco, CA, USA)	£170	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors: Yes	Steps, distance, Kcals, HR, active minutes, sleep	5 days	1	0.59%
Beurer AS80 (Beurer GmbH, Ulm, Germany)	£29.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	14 days	1	-58.07%
BodyMedia CORE (BodyMedia Inc., Pittsburgh	$150	Upper left arm	Research (commercially available)	Age, Gender, H, W	Accelerometer: Triaxial Heart rate:	Steps, activity intensity,	14 days	4	-1.06%

, PA, USA)					Heat sensors: Yes	Kcals, sleep			
Epson Pulsense (Epson, Suwa, Nagano Prefecture, Japan)	£79.99	Wrist	Commercial	Age, Gender, H, W, RHR	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, kcals, active minutes, HR, sleep	36 hours	1	-4.28%
ePulse Personal Fitness Assistant (ePulse) (Impact Sports Technologies, San Diego, CA, USA)	$129.95	Forearm	Commercial	Age, Gender, H, W, RHR	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Kcals, HR		1	-3.46%
Fitbit blaze (Fitbit Inc, San Francisco, California, USA	£134.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors: Triaxial accelerometer, altimeter, optical HR	Steps, distance, Kcals, active minutes, sleep, HR, steps	5 days	1	28.66%
Fitbit charge (Fitbit Inc, San Francisco, California, USA)	£109.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	5 days	2	-5.06%

					Triaxial accelerometer, altimeter				
Fitbit charge 2 (Fitbit Inc, San Francisco, California, USA)	£109.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, active minutes, sleep, HR, steps	5 days	1	-30.97%
Fitbit charge HR (Fitbit Inc, San Francisco, California, USA)	£139.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, active minutes, sleep, HR, steps	5 days	6	1.3%
Fitbit Flex (Fitbit Inc, San Francisco, California, USA)	£79.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	5 days	5	8.22%
Fitbit Surge (Fitbit Inc, San Francisco, California, USA)	£289.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, active minutes, altimeter, GPS	5 days		

Garmin Forerunner 225 (Garmin ltd, Olathe, Kansas, USA)	£199.99	Wrist	Commercial	Age, Gender, H, W, RHR, HRmax	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, HR, distance, Kcals, active minutes, altimeter, GPS	7-10 Hours	1	44.23%
Garmin Forerunner 920XT (Garmin ltd, Olathe, Kansas, USA)	£450	Wrist	Commercial	Age, Gender, H, W, RHR, HRmax	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, altimeter, sleep, HR, GPS	3 Days	1	-21.02%
Garmin vivoactive (Garmin ltd, Olathe, Kansas, USA)	£250	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, altimeter, sleep, GPS	7 Days	1	3.42%
Garmin vivofit (Garmin ltd, Olathe, Kansas, USA)	£79.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	1 Year	5	-26.09%
Garmin Vivosmart (Garmin ltd, Olathe, Kansas, USA)	£139.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	7 Days	1	5.98%

Garmin Vivosmart HR (Garmin ltd, Olathe, Kansas, USA)	£129.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, HR, intensity minutes, sleep	7 Days	1	16.85%
Jawbone UP (Jawbone, San Francisco, CA, USA)	£99.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Distance (app), Kcals, Steps, sleep	10 days	4	10.90%
Jawbone UP24 (Jawbone, San Francisco, CA, USA)	£89.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Distance (app), Kcal, Steps, sleep	14 Days	3	-29.58%
LifeChek calorie sensor (LifeChek, LLC, Pittsburgh, PA, USA)		Upper right arm	Commercial		Accelerometer: Triaxial Heart rate: Heat sensors: Yes	Kcals		1	-4.87%
Microsoft band (Microsoft Corporation, Redmond, WA, USA)	£169.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors: Yes	Steps, distance, kcals, active minutes, sleep, HR, GPS	48 Hours	1	-49.15%

Mio Alpha (Mio Global, Canada)	£119.99	Wrist	Commercial	Age, Gender H, W, HRMAX, RHR	Accelerometer: Heart rate: Yes Heat sensors:	Kcals, HR	24 Hours	1	-53.19%
Misfit Shine (Misfit, San Francisco, California, USA)	£99.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep		1	-2.36%
Nike Fuel Band (Nike Inc, Beaverton, OR, USA)	£129.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	4 days	3	-0.48%
Polar Loop (Polar Electro, Kempele, Finnland)	£49.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, distance, Kcals, active minutes, sleep	12 days	1	18.05%
Polar: AW200 (Polar Electro Oy, Kempele, Finland	€152 (watch +software)	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Heat sensors: Triaxial accelerometer	Steps, distance, Kcals, active minutes		1	-13.17%

Polar: AW360 (Polar Electro Oy, Kempele, Finland	£149.99	Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, active minutes, sleep, HR	12 Days	1	28.68%
Samsung Gear S (Samsung Electronics Co, Ltd, Suwon, South Korea)		Wrist	Commercial	Age, Gender, H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, Kcals, active minutes, sleep, HR, GPS	2 Days	1	-9.98%
SenseWear Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)	€800 (device)+ €1597 (software)	Upper right arm	Research	Age, Gender H, W,	Accelerometer: Biaxial Heart rate: Heat sensors: Yes	Steps, activity intensity, Kcals, sleep	14 days	12	-4.31%
SenseWear Mini Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)		Upper left arm	Research	Age, Gender H, W, smoking status	Accelerometer: Triaxial Heart rate: Heat sensors: Yes	Steps, activity intensity, Kcals, sleep	28 days	9	-1.44%
SenseWear Pro 2 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)		Upper right arm	Research	Age, Gender H, W, smoking status	Accelerometer: Biaxial Heart rate: Heat sensors: Yes	Steps, activity intensity, kcal, sleep	14 days	7	-7.54%

SenseWear Pro 3 Armband (HealthWear, Bodymedia, Pittsburgh, PA, USA)		Upper right arm	Research	Age, Gender H, W, smoking status	Accelerometer: Biaxial Heart rate: Heat sensors: Yes	Steps, activity intensity, kcal, sleep	14 days	12	-4.56%
TomTom Touch (TomTom, Amsterdam, the Netherlands)	£129.99	Wrist	Commercial	Age, Gender H, W	Accelerometer: Triaxial Heart rate: Yes Heat sensors:	Steps, distance, activity intensity, Kcal, sleep, HR,	5 Days	1	28.66%
Vivago (Vivago Wellness W, Paris, France).		Wrist	Commercial		Accelerometer: Triaxial Heart rate: Heat sensors:	Steps, activity intensity, Kcal, sleep		1	-8.02%
Withings Pulse (Withings, Issy-les-Moulineaux, France)	£39.99	Wrist, pocket or clip on	Commercial	Age, Gender H,	Accelerometer: Triaxial Heart rate: (non continuous) Heat sensors:	Steps, distance, Kcal, sleep	14 days	1	-133.33%
Withings Pulse 02 (Withings, Issy-les-Moulineaux, France)	£79.99	Wrist	Commercial	Age, Gender H, W	Accelerometer: Triaxial Heart rate: (non continuous) Heat sensors:	Steps, distance, activity intensity, Kcal, sleep,	14 days	2	-19.42%

HR, blood oxygen

S6:

	Activity	Hedges g	SE	Variance	Moderator	Moderator	Moderator
Alsubheen, 2016	AMBULATION	-1.26	0.32	0.10	GVF	ACC	Commercial
Bai, 2017	AMBULATION MODERATE	-0.12	0.15	0.02	AW	ACC + HR	Commercial
Bai, 2017	RUN MODERATE	-0.28	0.13	0.02	AW	ACC + HR	Commercial
Bai, 2017	SEDENTARY	-0.38	0.17	0.03	AW	ACC + HR	Commercial
Bai, 2017	AMBULATION MODERATE	0.72	0.13	0.02	FCHR	ACC + HR	Commercial
Bai, 2017	RUN MODERATE	0.37	0.13	0.02	FCHR	ACC + HR	Commercial
Bai, 2017	SEDENTARY	-0.56	0.18	0.03	FCHR	ACC + HR	Commercial
Benito, 2012	AEE LIGHT	-0.75	0.15	0.02	SWA p2	ACC + HS	Research
Benito, 2012	AEE MODERATE	-0.83	0.15	0.02	SWA p2	ACC + HS	Research
Benito, 2012	AEE VIGOROUS	-0.75	0.15	0.02	SWA p2	ACC + HS	Research
Berntsen, 2010	AEE MODERATE	0.27	0.16	0.03	SWA p2	ACC + HS	Research
Berntsen, 2010	AEE VERY VIGOROUS	-2.22	0.30	0.09	SWA p2	ACC + HS	Research
Berntsen, 2010	AEE VIGOROUS	-1.52	0.24	0.06	SWA p2	ACC + HS	Research
Berntsen, 2011	AEE LIGHT	-0.35	0.16	0.03	SWA p2	ACC + HS	Research
Berntsen, 2011	AEE MODERATE	-0.48	0.24	0.06	SWA p2	ACC + HS	Research
Bhammar, 2016	AEE	0.53	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	AEE 2	0.66	0.14	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	AMBULATION MODERATE	0.57	0.14	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	AMBULATION MODERATE 2	0.78	0.14	0.02	SWAM	ACC + HS	Research

Bhammar, 2016	AMBULATION VIGOROUS	0.31	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	AMBULATION VIGOROUS 2	0.58	0.14	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	BIKE LIGHT	-0.68	0.15	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	BIKE LIGHT 2	0.06	0.14	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	HOUSEHOLD	0.78	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	HOUSEHOLD 2	1.22	0.15	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	RUN MODERATE	0.82	0.14	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	RUN MODERATE 2	0.51	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	SEDENTARY	0.65	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	SEDENTARY 2	0.00	0.11	0.01	SWAM	ACC + HS	Research
Bhammar, 2016	SWEEP	0.65	0.13	0.02	SWAM	ACC + HS	Research
Bhammar, 2016	SWEEP 2	1.10	0.15	0.02	SWAM	ACC + HS	Research
Boudreaux, 2018	AEE	1.46	0.26	0.07	AWS2	ACC + HR	Commercial
Boudreaux, 2018	BIKE	1.70	0.27	0.07	AWS2	ACC + HR	Commercial
Boudreaux, 2018	AEE	-0.12	0.18	0.03	FB	ACC + HR	Commercial
Boudreaux, 2018	BIKE	-0.91	0.18	0.03	FB	ACC + HR	Commercial
Boudreaux, 2018	AEE	-0.20	0.19	0.04	FC2	ACC + HR	Commercial
Boudreaux, 2018	BIKE	-1.05	0.22	0.05	FC2	ACC + HR	Commercial
Boudreaux, 2018	AEE	0.65	0.20	0.04	GVHR	ACC + HR	Commercial
Boudreaux, 2018	BIKE	0.15	0.15	0.02	GVHR	ACC + HR	Commercial
Boudreaux, 2018	AEE	1.01	0.23	0.05	PA360	ACC + HR	Commercial
Boudreaux, 2018	BIKE	0.60	0.18	0.03	PA360	ACC + HR	Commercial
Boudreaux, 2018	AEE	0.57	0.19	0.04	TT	ACC + HR	Commercial
Boudreaux, 2018	BIKE	0.76	0.19	0.04	TT	ACC + HR	Commercial
Brazeau, 2011	BIKE LIGHT	-0.30	0.15	0.02	SWA p3	ACC + HS	Research
Brazeau, 2014	AMBULATION MODERATE B	1.30	0.22	0.05	SWA p3	ACC + HS	Research

Brazeau, 2014	AMBULATION MODERATE C	0.50	0.18	0.03	SWA p3	ACC + HS	Research
Brazeau, 2014	BIKE LIGHT B	-0.09	0.18	0.03	SWA p3	ACC + HS	Research
Brazeau, 2014	BIKE LIGHT C	-0.88	0.22	0.05	SWA p3	ACC + HS	Research
Brazeau, 2014	SEDENTARY B	2.16	0.27	0.07	SWA p3	ACC + HS	Research
Brazeau, 2014	SEDENTARY C	0.85	0.18	0.03	SWA p3	ACC + HS	Research
Brazeau, 2016	AMBULATION LIGHT	1.09	0.21	0.04	SWA p3	ACC + HS	Research
Brazeau, 2016	AMBULATION MODERATE	1.09	0.21	0.04	SWA p3	ACC + HS	Research
Brazeau, 2016	AMBULATION VIGOROUS	0.56	0.17	0.03	SWA p3	ACC + HS	Research
Brazeau, 2016	BIKE	-0.83	0.21	0.04	SWA p3	ACC + HS	Research
Brazeau, 2016	SEDENTARY	0.81	0.17	0.03	SWA p3	ACC + HS	Research
Brazeau, 2016	TEE	0.18	0.15	0.02	SWA p3	ACC + HS	Research
Brugniaux, 2010	AMBULATION VIGOROUS	-0.69	0.15	0.02	Polar AW200	ACC	Commercial
Calabro, 2014	AEE	0.14	0.11	0.01	SWA p3	ACC + HS	Research
Calabro, 2014	AEE	0.03	0.11	0.01	SWAM	ACC + HS	Research
Calabro, 2015	TEE	-0.04	0.12	0.02	SWAM	ACC + HS	Research
Casiraghi, 2013	BIKE LIGHT	-0.19	0.19	0.04	SWA	ACC + HS	Research
Choudhry, 2017	AMBULATION	0.33	0.17	0.03	AW	ACC + HR	Commercial
Choudhry, 2017	AMBULATION VIGOROUS	-0.05	0.17	0.03	AW	ACC + HR	Commercial
Choudhry, 2017	BIKE	-0.09	0.15	0.02	AW	ACC + HR	Commercial
Choudhry, 2017	COMPUTER	1.25	0.25	0.06	AW	ACC + HR	Commercial
Choudhry, 2017	HOUSEHOLD	-1.32	0.26	0.07	AW	ACC + HR	Commercial
Choudhry, 2017	RUN	0.59	0.16	0.03	AW	ACC + HR	Commercial
Choudhry, 2017	STAIRS	-1.00	0.21	0.04	AW	ACC + HR	Commercial
Choudhry, 2017	SWEEP	-1.14	0.24	0.06	AW	ACC + HR	Commercial
Choudhry, 2017	AMBULATION	0.34	0.14	0.02	BMC	ACC + HS	Research
Choudhry, 2017	AMBULATION VIGOROUS	0.52	0.14	0.02	BMC	ACC + HS	Research

Choudhry, 2017	BIKE	-0.53	0.16	0.02	BMC	ACC + HS	Research
Choudhry, 2017	COMPUTER	0.55	0.13	0.02	BMC	ACC + HS	Research
Choudhry, 2017	HOUSEHOLD	0.75	0.14	0.02	BMC	ACC + HS	Research
Choudhry, 2017	RUN	0.33	0.13	0.02	BMC	ACC + HS	Research
Choudhry, 2017	STAIRS	-1.01	0.17	0.03	BMC	ACC + HS	Research
Choudhry, 2017	SWEEP	0.93	0.15	0.02	BMC	ACC + HS	Research
Choudhry, 2017	AMBULATION	1.78	0.21	0.04	FCHR	ACC + HR	Commercial
Choudhry, 2017	AMBULATION VIGOROUS	1.60	0.19	0.04	FCHR	ACC + HR	Commercial
Choudhry, 2017	BIKE	-2.15	0.35	0.12	FCHR	ACC + HR	Commercial
Choudhry, 2017	COMPUTER	-0.24	0.20	0.04	FCHR	ACC + HR	Commercial
Choudhry, 2017	HOUSEHOLD	-0.29	0.20	0.04	FCHR	ACC + HR	Commercial
Choudhry, 2017	RUN	0.92	0.17	0.03	FCHR	ACC + HR	Commercial
Choudhry, 2017	STAIRS	-0.01	0.13	0.02	FCHR	ACC + HR	Commercial
Choudhry, 2017	SWEEP	0.89	0.23	0.05	FCHR	ACC + HR	Commercial
Choudhry, 2017	AMBULATION	0.91	0.17	0.03	JU24	ACC	Commercial
Choudhry, 2017	AMBULATION VIGOROUS	0.21	0.14	0.02	JU24	ACC	Commercial
Choudhry, 2017	BIKE	-3.67	0.40	0.16	JU24	ACC	Commercial
Choudhry, 2017	COMPUTER	-0.05	0.14	0.02	JU24	ACC	Commercial
Choudhry, 2017	HOUSEHOLD	-1.72	0.22	0.05	JU24	ACC	Commercial
Choudhry, 2017	RUN	0.92	0.17	0.03	JU24	ACC	Commercial
Choudhry, 2017	STAIRS	-1.39	0.20	0.04	JU24	ACC	Commercial
Choudhry, 2017	SWEEP	-1.05	0.18	0.03	JU24	ACC	Commercial
Choudhry, 2017	AMBULATION	-1.50	0.25	0.06	MB	ACC + HR + HS	Commercial
Choudhry, 2017	AMBULATION VIGOROUS	-2.15	0.31	0.10	MB	ACC + HR + HS	Commercial
Choudhry, 2017	BIKE	-0.90	0.22	0.05	MB	ACC + HR + HS	Commercial
Choudhry, 2017	COMPUTER	0.13	0.19	0.04	MB	ACC + HR + HS	Commercial
Choudhry, 2017	HOUSEHOLD	-2.13	0.35	0.12	MB	ACC + HR + HS	Commercial

Choudhry, 2017	RUN	0.17	0.17	0.03	MB	ACC + HR + HS	Commercial
Choudhry, 2017	STAIRS	-3.42	0.50	0.25	MB	ACC + HR + HS	Commercial
Choudhry, 2017	SWEEP	-2.01	0.33	0.11	MB	ACC + HR + HS	Commercial
Colbert, 2011	TEE	-1.10	0.11	0.01	SWA p3	ACC + HS	Research
Correa, 2016	AEE	0.56	0.11	0.01	ACT	ACC	Research
Correa, 2016	AEE	-0.43	0.11	0.01	SWA	ACC + HS	Research
Diaz, 2015	AMBULATION LIGHT L	-0.08	0.13	0.02	FF	ACC	Commercial
Diaz, 2015	AMBULATION LIGHT R	-0.15	0.13	0.02	FF	ACC	Commercial
Diaz, 2015	AMBULATION MODERATE L	0.95	0.16	0.02	FF	ACC	Commercial
Diaz, 2015	AMBULATION MODERATE R	0.96	0.16	0.02	FF	ACC	Commercial
Diaz, 2015	AMBULATION VIGOROUS L	1.44	0.19	0.03	FF	ACC	Commercial
Diaz, 2015	AMBULATION VIGOROUS R	0.94	0.15	0.02	FF	ACC	Commercial
Diaz, 2015	RUN L	0.57	0.14	0.02	FF	ACC	Commercial
Diaz, 2015	RUN R	0.27	0.13	0.02	FF	ACC	Commercial
Diaz, 2016	AMBULATION LIGHT	1.36	0.24	0.06	FF	ACC	Commercial
Diaz, 2016	AMBULATION MODERATE	3.04	0.41	0.17	FF	ACC	Commercial
Diaz, 2016	AMBULATION VIGOROUS	1.07	0.21	0.04	FF	ACC	Commercial
Diaz, 2016	RUN	0.78	0.19	0.04	FF	ACC	Commercial
Dondzilla, 2016	AMBULATION LIGHT	0.84	0.21	0.04	FC	ACC	Commercial
Dondzilla, 2016	AMBULATION MODERATE	-0.52	0.19	0.04	FC	ACC	Commercial
Dondzilla, 2016	RUN LIGHT	-0.51	0.19	0.04	FC	ACC	Commercial
Dondzilla, 2016	RUN MODERATE	-1.13	0.24	0.06	FC	ACC	Commercial

Dooley, 2017	AMBULATION LIGHT	0.49	0.13	0.02	AW	ACC + HR	Commercial
Dooley, 2017	AMBULATION MODERATE	0.29	0.12	0.02	AW	ACC + HR	Commercial
Dooley, 2017	AMBULATION VIGOROUS	0.27	0.12	0.01	AW	ACC + HR	Commercial
Dooley, 2017	SEDENTARY	1.48	0.19	0.04	AW	ACC + HR	Commercial
Dooley, 2017	STAND	1.80	0.21	0.05	AW	ACC + HR	Commercial
Dooley, 2017	AMBULATION LIGHT	1.60	0.13	0.02	FCHR	ACC + HR	Commercial
Dooley, 2017	AMBULATION MODERATE	1.13	0.11	0.01	FCHR	ACC + HR	Commercial
Dooley, 2017	AMBULATION VIGOROUS	0.08	0.09	0.01	FCHR	ACC + HR	Commercial
Dooley, 2017	SEDENTARY	0.28	0.14	0.02	FCHR	ACC + HR	Commercial
Dooley, 2017	STAND	-0.39	0.14	0.02	FCHR	ACC + HR	Commercial
Dooley, 2017	AMBULATION LIGHT	0.72	0.10	0.01	GF225	ACC + HR	Commercial
Dooley, 2017	AMBULATION MODERATE	1.27	0.12	0.01	GF225	ACC + HR	Commercial
Dooley, 2017	AMBULATION VIGOROUS	0.63	0.10	0.01	GF225	ACC + HR	Commercial
Dooley, 2017	SEDENTARY	1.45	0.20	0.04	GF225	ACC + HR	Commercial
Dooley, 2017	STAND	1.14	0.18	0.03	GF225	ACC + HR	Commercial
Drenowatz, 2011	RUN	-2.21	0.30	0.09	SWA	ACC + HS	Research
Drenowatz, 2011	RUN LIGHT	-1.10	0.20	0.04	SWA	ACC + HS	Research
Drenowatz, 2011	RUN MODERATE	-1.95	0.28	0.08	SWA	ACC + HS	Research
Drenowatz, 2011	RUN VIGOROUS	-2.41	0.32	0.10	SWA	ACC + HS	Research
Erdogan, 2010	AEE	0.08	0.11	0.01	SWA	ACC + HS	Research
Erdogan, 2010	AEE VIGOROUS	-0.01	0.11	0.01	SWA	ACC + HS	Research
Fruin, 2004	AMBULATION MODERATE	0.23	0.16	0.03	SWA	ACC + HS	Research
Fruin, 2004	AMBULATION	-0.21	0.16	0.03	SWA	ACC + HS	Research

	UPHILL VIGOROUS						
Fruin, 2004	AMBULATION VIGOROUS	0.51	0.17	0.03	SWA	ACC + HS	Research
Fruin, 2004	BIKE EARLY	-0.50	0.23	0.05	SWA	ACC + HS	Research
Fruin, 2004	BIKE LATE	-0.05	0.21	0.05	SWA	ACC + HS	Research
Fruin, 2004	BIKE MIDDLE	-0.27	0.22	0.05	SWA	ACC + HS	Research
Furlanetto, 2010	AMBULATION LIGHT	-0.19	0.13	0.02	SWA	ACC + HS	Research
Furlanetto, 2010	AMBULATION MODERATE	-0.04	0.13	0.02	SWA	ACC + HS	Research
Furlanetto, 2010	AMBULATION VIGOROUS	-0.11	0.13	0.02	SWA	ACC + HS	Research
Gastin, 2017	AEE 1	-1.71	0.22	0.05	SWA	ACC + HS	Research
Gastin, 2017	AEE 2	-1.83	0.23	0.05	SWA	ACC + HS	Research
Gastin, 2017	AEE 3	-1.61	0.22	0.05	SWA	ACC + HS	Research
Gastin, 2017	AMBULATION	2.87	0.33	0.11	SWA	ACC + HS	Research
Gastin, 2017	RUN LIGHT	1.06	0.18	0.03	SWA	ACC + HS	Research
Gastin, 2017	RUN MODERATE	-0.68	0.16	0.02	SWA	ACC + HS	Research
Heiermann, 2011	REST	0.76	0.13	0.02	SWA p2	ACC + HS	Research
Imboden, 2017	AEE	-0.65	0.12	0.02	FF	ACC	Commercial
Imboden, 2017	AEE	-1.30	0.15	0.02	JU24	ACC	Commercial
Jakicic, 2004	AEE	2.43	0.25	0.06	SWA	ACC + HS	Research
Jakicic, 2004	AEE 1	0.90	0.14	0.02	SWA	ACC + HS	Research
Jakicic, 2004	AMBULATION MODERATE	1.92	0.22	0.05	SWA	ACC + HS	Research
Jakicic, 2004	AMBULATION UPHILL MODERATE	-0.31	0.13	0.02	SWA	ACC + HS	Research
Jakicic, 2004	AMBULATION UPHILL VIGOROUS	-1.58	0.20	0.04	SWA	ACC + HS	Research
Jakicic, 2004	BIKE LIGHT	-0.41	0.14	0.02	SWA	ACC + HS	Research
Jakicic, 2004	BIKE MODERATE	-2.28	0.27	0.07	SWA	ACC + HS	Research
Jakicic, 2004	STAIRS LIGHT	-0.18	0.14	0.02	SWA	ACC + HS	Research

Jakicic, 2004	STAIRS MODERATE	-1.46	0.20	0.04	SWA	ACC + HS	Research
Johannsen, 2010	TEE	-0.20	0.12	0.01	SWA p3	ACC + HS	Research
Johannsen, 2010	TEE	-0.04	0.12	0.01	SWAM	ACC + HS	Research
Kim, 2015	AEE LIGHT	0.16	0.10	0.01	BMC	ACC + HS	Research
Kim, 2015	AEE MODERATE	-0.09	0.10	0.01	BMC	ACC + HS	Research
Kim, 2015	AEE VIGOROUS	0.29	0.10	0.01	BMC	ACC + HS	Research
Kim, 2015	SEDENTARY	0.30	0.09	0.01	BMC	ACC + HS	Research
King, 2004	AMBULATION LIGHT M	4.46	0.78	0.60	SWA	ACC + HS	Research
King, 2004	AMBULATION M	1.56	0.34	0.11	SWA	ACC + HS	Research
King, 2004	AMBULATION MODERATE F	1.94	0.37	0.14	SWA	ACC + HS	Research
King, 2004	AMBULATION MODERATE M	3.76	0.66	0.44	SWA	ACC + HS	Research
King, 2004	AMBULATION VIGOROUS F	2.05	0.39	0.15	SWA	ACC + HS	Research
King, 2004	AMBULETION F	0.72	0.24	0.06	SWA	ACC + HS	Research
King, 2004	RUN LIGHT F	2.20	0.45	0.20	SWA	ACC + HS	Research
King, 2004	RUN LIGHT M	2.73	0.56	0.31	SWA	ACC + HS	Research
King, 2004	RUN MODERATE F	1.77	0.39	0.15	SWA	ACC + HS	Research
King, 2004	RUN MODERATE M	1.69	0.39	0.15	SWA	ACC + HS	Research
King, 2004	RUN VERY VIGOROUS F	0.71	0.26	0.07	SWA	ACC + HS	Research
King, 2004	RUN VERY VIGOROUS M	0.69	0.27	0.07	SWA	ACC + HS	Research
King, 2004	RUN VIGOROUS F	1.32	0.33	0.11	SWA	ACC + HS	Research
King, 2004	RUN VIGOROUS M	0.48	0.25	0.06	SWA	ACC + HS	Research
Koehler, 2011	TEE	-0.06	0.17	0.03	SWA p3	ACC + HS	Research
Lee, 2011	AMBULATION LIGHT	-0.74	0.20	0.04	EPUL	ACC + HR	Commercial

Lee, 2011	AMBULATION MODERATE	0.04	0.18	0.03	EPUL	ACC + HR	Commercial
Lee, 2011	RUN LIGHT	0.01	0.18	0.03	EPUL	ACC + HR	Commercial
Lee, 2011	RUN MODERATE	0.12	0.18	0.03	EPUL	ACC + HR	Commercial
Lee, 2011	SEDENTARY	-0.31	0.18	0.03	EPUL	ACC + HR	Commercial
Lee, 2014	AEE	-1.37	0.18	0.03	BB1	ACC + HR + HS	Commercial
Lee, 2014	AEE	-0.28	0.10	0.01	BMC	ACC + HS	Research
Lee, 2014	AEE	-0.34	0.08	0.01	JU	ACC	Commercial
Lee, 2014	AEE	-0.10	0.09	0.01	NF	ACC	Commercial
Mackey, 2011	TEE	0.05	0.15	0.02	SWA	ACC + HS	Research
Mackey, 2011	TEE 6	-0.06	0.15	0.02	SWA	ACC + HS	Research
Martien, 2015	AEE	-0.25	0.11	0.01	SWAM	ACC + HS	Research
Martien, 2015	SEDENTARY	-0.99	0.11	0.01	SWAM	ACC + HS	Research
McMinn, 2013	AMBULATION LIGHT	-1.55	0.25	0.06	AGT3X	ACC	Research
McMinn, 2013	AMBULATION MODERATE	0.04	0.16	0.03	AGT3X	ACC	Research
McMinn, 2013	AMBULATION VIGOROUS	0.48	0.17	0.03	AGT3X	ACC	Research
Melanson, 2009	TEE	-0.05	0.35	0.12	LC	ACC + HS	Commercial
Mikulic, 2011	AEE	-1.95	0.28	0.08	SWA p3	ACC + HS	Research
Montoye, 2017	AMBULATION LIGHT	0.23	0.12	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	AMBULATION MODERATE	0.22	0.12	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	AMBULATION UPHILL	0.42	0.13	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	AMBULATION UPHILL LIGHT	0.45	0.13	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	AMBULATION UPHILL MODERATE	0.47	0.13	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	AMBULATION VIGOROUS	0.36	0.13	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	BIKE	0.43	0.18	0.03	FCHR	ACC + HR	Commercial

Montoye, 2017	RUN	0.52	0.15	0.02	FCHR	ACC + HR	Commercial
Montoye, 2017	SITTING	0.06	0.19	0.04	FCHR	ACC + HR	Commercial
Montoye, 2017	STAND	0.00	0.19	0.04	FCHR	ACC + HR	Commercial
Montoye, 2017	SUPINE	-0.23	0.19	0.04	FCHR	ACC + HR	Commercial
Murakami, 2016	TEE	-0.68	0.19	0.04	EP	ACC + HR	Commercial
Murakami, 2016	TEE MC	0.12	0.17	0.03	EP	ACC + HR	Commercial
Murakami, 2016	TEE	-0.48	0.15	0.02	FF	ACC	Commercial
Murakami, 2016	TEE MC	0.38	0.11	0.01	FF	ACC	Commercial
Murakami, 2016	TEE	-1.62	0.30	0.09	GVF	ACC	Commercial
Murakami, 2016	TEE MC	-0.83	0.23	0.05	GVF	ACC	Commercial
Murakami, 2016	TEE	-1.96	0.30	0.09	JU24	ACC	Commercial
Murakami, 2016	TEE MC	-0.95	0.21	0.04	JU24	ACC	Commercial
Murakami, 2016	TEE	-0.68	0.27	0.07	MS	ACC	Commercial
Murakami, 2016	TEE MC	0.40	0.25	0.06	MS	ACC	Commercial
Murakami, 2016	TEE	-1.68	0.23	0.05	WPO	ACC	Commercial
Murakami, 2016	TEE MC	-0.93	0.18	0.03	WPO	ACC	Commercial
Nelson, 2016	AMBULATION	0.92	0.14	0.02	FF	ACC	Commercial
Nelson, 2016	HOUSEHOLD	-0.27	0.14	0.02	FF	ACC	Commercial
Nelson, 2016	SEDENTARY	-0.33	0.14	0.02	FF	ACC	Commercial
Nelson, 2016	AMBULATION	0.48	0.12	0.01	JU	ACC	Commercial
Nelson, 2016	HOUSEHOLD	-1.30	0.24	0.06	JU	ACC	Commercial
Nelson, 2016	SEDENTARY	-0.39	0.18	0.03	JU	ACC	Commercial
Papazoglou, 2006	AMBULATION	0.82	0.19	0.04	SWA p2	ACC + HS	Research
Papazoglou, 2006	BIKE	0.54	0.17	0.03	SWA p2	ACC + HS	Research
Papazoglou, 2006	STAIRS	0.88	0.17	0.03	SWA p2	ACC + HS	Research
Price, 2017	AMBULATION	-1.15	0.29	0.09	GVF	ACC	Commercial
Price, 2017	RUN	-0.50	0.31	0.09	GVF	ACC	Commercial
Price, 2017	AMBULATION	-0.12	0.16	0.03	JU	ACC	Commercial

Price, 2017	RUN	0.48	0.19	0.04	JU	ACC	Commercial
Reece, 2015	AMBULATION	0.40	0.16	0.03	SWAM	ACC + HS	Research
Reece, 2015	COMPUTER	-1.22	0.19	0.04	SWAM	ACC + HS	Research
Reece, 2015	SEDENTARY	0.20	0.14	0.02	SWAM	ACC + HS	Research
Reece, 2015	SITTING	0.19	0.14	0.02	SWAM	ACC + HS	Research
Reece, 2015	STAND	-0.58	0.15	0.02	SWAM	ACC + HS	Research
Reece, 2015	STAND COMPUTER	-1.10	0.18	0.03	SWAM	ACC + HS	Research
Rousset, 2015	TEE CC	-0.49	0.10	0.01	SWA p3	ACC + HS	Research
Rousset, 2015	TEE DLW	-0.05	0.10	0.01	SWA p3	ACC + HS	Research
Ryan, 2013	AMBULATION	0.82	0.17	0.03	SWA p2	ACC + HS	Research
Ryan, 2013	AMBULATION MODERATE	-0.08	0.14	0.02	SWA p2	ACC + HS	Research
Ryan, 2013	AMBULATION UPHILL MODERATE	-2.27	0.27	0.08	SWA p2	ACC + HS	Research
Ryan, 2013	RUN	-0.38	0.15	0.02	SWA p2	ACC + HS	Research
Ryan, 2013	SEDENTARY	0.65	0.14	0.02	SWA p2	ACC + HS	Research
Slinde, 2013	TEE	0.20	0.09	0.01	SWA p2	ACC + HS	Research
Slinde, 2013	TEE 6	-0.67	0.09	0.01	SWA p2	ACC + HS	Research
Smith, 2012	AMBULATION	1.29	0.18	0.03	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION 2	1.96	0.23	0.05	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION LIGHT	1.43	0.19	0.04	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION LIGHT 2	1.77	0.22	0.05	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION MODERATE	1.57	0.20	0.04	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION MODERATE 2	1.43	0.19	0.04	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION UPHILL MODERATE	-0.16	0.13	0.02	SWAM	ACC + HS	Research
Smith, 2012	AMBULATION UPHILL MODERATE 2	-0.22	0.13	0.02	SWAM	ACC + HS	Research

Smith, 2012	COMPUTER	0.37	0.12	0.02	SWAM	ACC + HS	Research
Smith, 2012	HOUSEHOLD	1.25	0.16	0.03	SWAM	ACC + HS	Research
Smith, 2012	HOUSEHOLD 2	1.78	0.20	0.04	SWAM	ACC + HS	Research
Smith, 2012	SEDENTARY 2	0.36	0.12	0.02	SWAM	ACC + HS	Research
Smith, 2012	SWEEP	0.52	0.13	0.02	SWAM	ACC + HS	Research
Smith, 2012	SWEEP 2	1.50	0.18	0.03	SWAM	ACC + HS	Research
Stackpool, 2015	AEE	-0.77	0.18	0.03	BMC	ACC + HS	Research
Stackpool, 2015	AEE 1	-1.29	0.22	0.05	BMC	ACC + HS	Research
Stackpool, 2015	AMBULATION	0.16	0.17	0.03	BMC	ACC + HS	Research
Stackpool, 2015	RUN	-0.65	0.18	0.03	BMC	ACC + HS	Research
Stackpool, 2015	AEE	-1.15	0.17	0.03	JU	ACC	Commercial
Stackpool, 2015	AEE 1	0.00	0.13	0.02	JU	ACC	Commercial
Stackpool, 2015	AMBULATION	0.56	0.15	0.02	JU	ACC	Commercial
Stackpool, 2015	RUN	0.77	0.19	0.03	JU	ACC	Commercial
Stackpool, 2015	AEE	-0.63	0.16	0.03	NF	ACC	Commercial
Stackpool, 2015	AEE 1	-1.14	0.19	0.04	NF	ACC	Commercial
Stackpool, 2015	AMBULATION	-0.08	0.15	0.02	NF	ACC	Commercial
Stackpool, 2015	RUN	0.63	0.17	0.03	NF	ACC	Commercial
St-Onge, 2007	TEE	0.27	0.10	0.01	SWA	ACC + HS	Research
Tucker, 2015	AEE	0.04	0.13	0.02	NF	ACC	Commercial
Tucker, 2015	AEE	-0.08	0.15	0.02	SWA	ACC + HS	Research
Van Helst, 2012	AMBULATION	-1.42	0.20	0.04	V	ACC	Commercial
Van Helst, 2012	RUN MODERATE	-0.22	0.14	0.02	V	ACC	Commercial
Van Helst, 2012	RUN VIGOROUS	-0.23	0.14	0.02	V	ACC	Commercial
Van Helst, 2012	SEDENTARY	0.00	0.14	0.02	V	ACC	Commercial
Van Hoye, 2014	AMBULATION F	0.14	0.16	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION LIGHT F	-0.52	0.17	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION LIGHT M	-1.06	0.19	0.04	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION M	-0.08	0.15	0.02	SWA p3	ACC + HS	Research

Van Hoye, 2014	AMBULATION MODERATE F	-0.47	0.17	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION MODERATE M	-1.13	0.20	0.04	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION VIGOROUS F	-0.62	0.18	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	AMBULATION VIGOROUS M	-1.16	0.20	0.04	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN LIGHT F	-1.30	0.22	0.05	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN LIGHT M	-1.35	0.21	0.04	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN MODERATE F	-1.88	0.28	0.08	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN MODERATE M	-2.02	0.26	0.07	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN VERY LIGHT F	-0.76	0.18	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN VERY LIGHTM	-0.93	0.18	0.03	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN VERY VIGOROUS M	-3.08	0.41	0.17	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN VIGOROUS F	-2.24	0.33	0.11	SWA p3	ACC + HS	Research
Van Hoye, 2014	RUN VIGOROUS M	-3.03	0.36	0.13	SWA p3	ACC + HS	Research
Van Hoye, 2015	AMBULATION LIGHT	-0.36	0.12	0.01	SWA p3	ACC + HS	Research
Van Hoye, 2015	AMBULATION MODERATE	-0.28	0.12	0.01	SWA p3	ACC + HS	Research
Van Hoye, 2015	RUN	-1.04	0.14	0.02	SWA p3	ACC + HS	Research
Van Hoye, 2015	RUN MODERATE	-0.77	0.13	0.02	SWA p3	ACC + HS	Research
Van Hoye, 2015	STAND	0.23	0.11	0.01	SWA p3	ACC + HS	Research
Van Hoye, 2015	STAND 1	0.62	0.12	0.01	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION	0.78	0.19	0.03	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION DOWNHILL LIGHT	2.01	0.29	0.08	SWA p3	ACC + HS	Research

Vernillo, 2015	AMBULATION DOWNHILL MODERATE	2.28	0.31	0.10	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION DOWNHILL VIGOROUS	2.09	0.29	0.09	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION MODERATE	0.30	0.16	0.03	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION UPHILL VIGOROUS	-2.33	0.32	0.10	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION UPHILL MODERATE	-1.07	0.20	0.04	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION UPHILL VERY VIGOROUS	-2.83	0.37	0.14	SWA p3	ACC + HS	Research
Vernillo, 2015	SEDENTARY	0.00	0.15	0.02	SWA p3	ACC + HS	Research
Vernillo, 2015	AMBULATION	0.56	0.17	0.03	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION DOWNHILL LIGHT	2.29	0.32	0.10	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION DOWNHILL MODERATE	2.28	0.31	0.10	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION DOWNHILL VIGOROUS	2.27	0.31	0.10	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION MODERATE	0.00	0.16	0.03	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION UPHILL MODERATE	-1.49	0.24	0.06	SWAM	ACC + HS	Research
Vernillo, 2015	AMBULATION UPHILL	-2.78	0.37	0.13	SWAM	ACC + HS	Research

Vernillo, 2015	VERY VIGOROUS AMBULATION UPHILL VIGOROUS	-2.46	0.33	0.11	SWAM	ACC + HS	Research
Vernillo, 2015	SEDENTARY	0.00	0.15	0.02	SWAM	ACC + HS	Research
Wahl, 2017	AMBULATION	0.14	0.22	0.05	BA	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	-1.13	0.28	0.08	BA	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.40	0.22	0.05	BA	ACC	Commercial
Wahl, 2017	RUN MODERATE	-0.68	0.24	0.06	BA	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.83	0.34	0.12	BA	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	-0.82	0.25	0.06	BA	ACC	Commercial
Wahl, 2017	AMBULATION	1.12	0.28	0.08	FC	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	0.00	0.21	0.05	FC	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.58	0.23	0.05	FC	ACC	Commercial
Wahl, 2017	RUN MODERATE	0.37	0.22	0.05	FC	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.19	0.29	0.09	FC	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	-0.05	0.21	0.05	FC	ACC	Commercial
Wahl, 2017	AMBULATION	0.78	0.25	0.06	FCHR	ACC + HR	Commercial
Wahl, 2017	RUN INTERMITTENT	0.42	0.22	0.05	FCHR	ACC + HR	Commercial
Wahl, 2017	RUN LIGHT	0.11	0.22	0.05	FCHR	ACC + HR	Commercial
Wahl, 2017	RUN MODERATE	0.33	0.22	0.05	FCHR	ACC + HR	Commercial
Wahl, 2017	RUN OUTDOOR	-0.35	0.30	0.09	FCHR	ACC + HR	Commercial
Wahl, 2017	RUN VIGOROUS	0.31	0.22	0.05	FCHR	ACC + HR	Commercial
Wahl, 2017	AMBULATION	-0.50	0.23	0.05	GF920XT	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	-0.16	0.22	0.05	GF920XT	ACC	Commercial
Wahl, 2017	RUN LIGHT	-0.32	0.22	0.05	GF920XT	ACC	Commercial

Wahl, 2017	RUN MODERATE	-0.19	0.22	0.05	GF920XT	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.47	0.31	0.09	GF920XT	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	-0.22	0.22	0.05	GF920XT	ACC	Commercial
Wahl, 2017	AMBULATION	-0.16	0.22	0.05	GVA	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	-0.01	0.21	0.05	GVA	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.46	0.23	0.05	GVA	ACC	Commercial
Wahl, 2017	RUN MODERATE	0.95	0.26	0.07	GVA	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.25	0.29	0.09	GVA	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	0.15	0.22	0.05	GVA	ACC	Commercial
Wahl, 2017	AMBULATION	-0.13	0.22	0.05	GVF	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	-0.50	0.23	0.05	GVF	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.22	0.22	0.05	GVF	ACC	Commercial
Wahl, 2017	RUN MODERATE	0.15	0.22	0.05	GVF	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.90	0.35	0.12	GVF	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	-0.29	0.22	0.05	GVF	ACC	Commercial
Wahl, 2017	AMBULATION	-0.10	0.22	0.05	GVS	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	0.03	0.21	0.05	GVS	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.37	0.22	0.05	GVS	ACC	Commercial
Wahl, 2017	RUN MODERATE	0.39	0.22	0.05	GVS	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.05	0.29	0.08	GVS	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	0.14	0.22	0.05	GVS	ACC	Commercial
Wahl, 2017	AMBULATION	0.36	0.22	0.05	PL	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	0.02	0.21	0.05	PL	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.27	0.22	0.05	PL	ACC	Commercial

Wahl, 2017	RUN MODERATE	0.34	0.22	0.05	PL	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	0.15	0.22	0.05	PL	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	0.25	0.22	0.05	PL	ACC	Commercial
Wahl, 2017	AMBULATION	-0.11	0.22	0.05	SWAM	ACC + HS	Research
Wahl, 2017	RUN INTERMITTENT	-0.43	0.23	0.05	SWAM	ACC + HS	Research
Wahl, 2017	RUN LIGHT	-0.10	0.22	0.05	SWAM	ACC + HS	Research
Wahl, 2017	RUN MODERATE	-0.59	0.23	0.05	SWAM	ACC + HS	Research
Wahl, 2017	RUN OUTDOOR	-0.70	0.33	0.11	SWAM	ACC + HS	Research
Wahl, 2017	RUN VIGOROUS	-0.85	0.25	0.06	SWAM	ACC + HS	Research
Wahl, 2017	AMBULATION	-1.12	0.38	0.15	WPO	ACC	Commercial
Wahl, 2017	RUN INTERMITTENT	-1.86	0.51	0.26	WPO	ACC	Commercial
Wahl, 2017	RUN LIGHT	0.12	0.29	0.08	WPO	ACC	Commercial
Wahl, 2017	RUN MODERATE	0.04	0.29	0.08	WPO	ACC	Commercial
Wahl, 2017	RUN OUTDOOR	-0.15	0.29	0.08	WPO	ACC	Commercial
Wahl, 2017	RUN VIGOROUS	-0.27	0.30	0.09	WPO	ACC	Commercial
Wallen 2016	AEE	-2.44	0.27	0.07	AW	ACC + HR	Commercial
Wallen 2016	AEE	-0.80	0.16	0.03	FCHR	ACC + HR	Commercial
Wallen 2016	AEE	-1.19	0.28	0.08	MA	HR	Commercial
Wallen 2016	AEE	-0.54	0.16	0.03	SG	ACC + HR	Commercial
Woodman, 2017	AMBULATION	2.25	0.34	0.12	BP	ACC + HR + HS	Commercial
Woodman, 2017	AMBULATION UPHILL MODERATE	0.73	0.20	0.04	BP	ACC + HR + HS	Commercial
Woodman, 2017	BIKE LIGHT	-0.45	0.20	0.04	BP	ACC + HR + HS	Commercial
Woodman, 2017	BIKE MODERATE	-1.01	0.23	0.05	BP	ACC + HR + HS	Commercial
Woodman, 2017	COMPUTER	-0.37	0.20	0.04	BP	ACC + HR + HS	Commercial

Woodman, 2017	HOUSEHOLD	-0.78	0.22	0.05	BP	ACC + HR + HS	Commercial
Woodman, 2017	RUN	0.20	0.16	0.03	BP	ACC + HR + HS	Commercial
Woodman, 2017	SEATED	0.03	0.19	0.04	BP	ACC + HR + HS	Commercial
Woodman, 2017	SEDENTARY	-0.97	0.23	0.05	BP	ACC + HR + HS	Commercial
Woodman, 2017	STAIRS	0.49	0.19	0.04	BP	ACC + HR + HS	Commercial
Woodman, 2017	SWEEP	-1.80	0.31	0.10	BP	ACC + HR + HS	Commercial
Woodman, 2017	AMBULATION	-1.36	0.23	0.05	GVF	ACC	Commercial
Woodman, 2017	AMBULATION UPHILL MODERATE	-2.84	0.38	0.14	GVF	ACC	Commercial
Woodman, 2017	BIKE LIGHT	-6.59	0.84	0.71	GVF	ACC	Commercial
Woodman, 2017	BIKE MODERATE	-0.80	0.19	0.04	GVF	ACC	Commercial
Woodman, 2017	COMPUTER	-0.27	0.17	0.03	GVF	ACC	Commercial
Woodman, 2017	HOUSEHOLD	-2.31	0.32	0.10	GVF	ACC	Commercial
Woodman, 2017	RUN	-0.98	0.28	0.08	GVF	ACC	Commercial
Woodman, 2017	SEATED	-1.21	0.22	0.05	GVF	ACC	Commercial
Woodman, 2017	SEDENTARY	-0.32	0.17	0.03	GVF	ACC	Commercial
Woodman, 2017	STAIRS	-4.16	0.53	0.28	GVF	ACC	Commercial
Woodman, 2017	SWEEP	-2.18	0.31	0.10	GVF	ACC	Commercial
Woodman, 2017	AMBULATION	-1.88	0.24	0.06	WP	ACC	Commercial
Woodman, 2017	AMBULATION UPHILL MODERATE	-2.80	0.32	0.10	WP	ACC	Commercial
Woodman, 2017	BIKE LIGHT	-5.53	0.61	0.38	WP	ACC	Commercial
Woodman, 2017	BIKE MODERATE	-2.25	0.28	0.08	WP	ACC	Commercial
Woodman, 2017	COMPUTER	1.94	0.25	0.06	WP	ACC	Commercial
Woodman, 2017	HOUSEHOLD	-0.83	0.17	0.03	WP	ACC	Commercial
Woodman, 2017	RUN	-2.37	0.30	0.09	WP	ACC	Commercial
Woodman, 2017	SEATED	-1.41	0.20	0.04	WP	ACC	Commercial

Woodman, 2017	SEDENTARY	-1.05	0.18	0.03	WP	ACC	Commercial
Woodman, 2017	STAIRS	-3.47	0.39	0.15	WP	ACC	Commercial
Woodman, 2017	SWEEP	-1.76	0.23	0.05	WP	ACC	Commercial

S7:

			Heterogeneity		Effect size				Publication bias			
Overall activities												
		n	I-squared (between studies)	P-value	Hedges' g (95% CI)	Lower Limit	Upper limit	P-value	Egger's intercept	Lower limit	Upper limit	P-Value
	ACT	1.00	0.00	1.00	0.56	-0.46	1.58	0.28				
	AGT3X	1.00	0.00	1.00	-0.35	-1.42	0.73	0.53				
	AW	4.00	97.30	0.00	-0.43	-0.97	0.10	0.11	-19.41	-65.76	26.94	0.21
	AWS2	1.00	0.00	1.00	1.58	0.45	2.70	0.01				
	BA	1.00	0.00	1.00	-0.49	-1.61	0.64	0.40				
	BB1	1.00	0.00	1.00	-1.37	-2.42	-0.31	0.01				
	BMC	4.00	87.47	0.00	-0.12	-0.64	0.40	0.65	-2.60	-30.95	25.74	0.73
	BP	1.00	0.00	1.00	-0.15	-1.25	0.94	0.78				
	EP	1.00	0.00	1.00	-0.28	-1.34	0.78	0.60				
	EPUL	1.00	0.00	1.00	-0.18	-1.24	0.88	0.74				
	FB	1.00	0.00	1.00	-0.51	-1.57	0.54	0.34				
	FC	2.00	74.54	0.05	-0.02	-0.79	0.75	0.95				
	FC2	1.00	0.00	1.00	-0.62	-1.70	0.45	0.26				
	FCHR	6.00	89.06	0.00	0.13	-0.31	0.56	0.57	-2.32	-20.26	15.62	0.74
	FF	5.00	94.80	0.00	0.27	-0.20	0.74	0.26	13.81	-4.45	32.07	0.09
	GF225	1.00	0.00	1.00	1.04	0.00	2.08	0.05				
	GF920XT	1.00	0.00	1.00	-0.31	-1.41	0.79	0.58				

GVA	1.00	0.00	1.00	0.19	-0.91	1.29	0.74				
GVF	5.00	79.33	0.00	-1.09	-1.60	-0.57	0.00	-11.66	-24.75	1.42	0.06
GVHR	1.00	0.00	1.00	0.40	-0.66	1.46	0.46				
GVS	1.00	0.00	1.00	0.13	-0.97	1.23	0.81				
JU	4.00	73.04	0.01	-0.13	-0.66	0.39	0.62	2.81	-7.67	13.28	0.37
JU24	3.00	66.91	0.05	-1.16	-1.78	-0.54	0.00	1.15	-71.75	74.05	0.87
LC	1.00	0.00	1.00	-0.05	-1.26	1.15	0.93				
MA	1.00	0.00	1.00	-1.19	-2.34	-0.05	0.04				
MB	1.00	0.00	1.00	-1.48	-2.64	-0.31	0.01				
MS	1.00	0.00	1.00	-0.14	-1.26	0.98	0.80				
NF	3.00	25.44	0.26	-0.12	-0.72	0.48	0.69	-1.04	-43.09	41.00	0.80
PA360	1.00	0.00	1.00	0.80	-0.27	1.88	0.14				
PL	1.00	0.00	1.00	0.23	-0.85	1.32	0.67				
Polar AW200	1.00	0.00	1.00	-0.69	-1.73	0.35	0.19				
SG	1.00	0.00	1.00	-0.54	-1.59	0.51	0.31				
SWA	12.00	87.57	0.00	-0.12	-0.43	0.19	0.45	-1.11	-6.72	4.49	0.67
SWA p2	7.00	94.47	0.00	-0.17	-0.57	0.22	0.39	-2.05	-16.37	12.26	0.73
SWA p3	12.00	93.03	0.00	-0.32	-0.62	-0.01	0.04	-0.49	-8.81	7.82	0.89
SWAM	9.00	91.19	0.00	0.02	-0.33	0.37	0.90	2.30	-8.15	12.76	0.61
TT	1.00	0.00	1.00	0.67	-0.40	1.73	0.22				
V	1.00	0.00	1.00	-0.47	-1.51	0.57	0.38				
WP	1.00	0.00	1.00	-1.95	-3.12	-0.78	0.00				
WPO	2.00	71.58	0.06	-0.97	-1.77	-0.16	0.02				
Between			0.00								

	Overall	104.00	92.18		-0.23	-0.44	-0.03	0.03				
AEE		n	I-squared (between studies)	P-value	Effect size (Hedges' g) (95% CI)	Lower Limit	Upper limit	P-value				
	ACT	1.00	0.00	1.00	0.56	-0.61	1.73	0.35				
	AW	1.00	0.00	1.00	-2.44	-3.71	-1.18	0.00				
	AWS2	1.00	0.00	1.00	1.46	0.21	2.71	0.02				
	BB1	1.00	0.00	1.00	-1.37	-2.57	-0.17	0.03				
	BMC	3.00	92.83	0.00	-0.38	-1.06	0.30	0.28	-8.62	-94.18	76.94	0.42
	FB	1.00	0.00	1.00	-0.12	-1.32	1.08	0.85				
	FC2	1.00	0.00	1.00	-0.20	-1.40	1.01	0.75				
	FCHR	1.00	0.00	1.00	-0.80	-1.99	0.39	0.19				
	FF	1.00	0.00	1.00	-0.65	-1.82	0.52	0.28				
	GVHR	1.00	0.00	1.00	0.65	-0.56	1.86	0.29				
	JU	2.00	47.50	0.17	-0.46	-1.28	0.37	0.28				
	JU24	1.00	0.00	1.00	-1.30	-2.48	-0.12	0.03				
	MA	1.00	0.00	1.00	-1.19	-2.47	0.08	0.07				
	NF	3.00	89.88	0.00	-0.31	-0.99	0.37	0.38	-5.94	-91.55	99.65	0.53
	PA360	1.00	0.00	1.00	1.01	-0.22	2.24	0.11				
	SG	1.00	0.00	1.00	-0.54	-1.73	0.65	0.38				
	SWA	5.00	97.12	0.00	-0.10	-0.63	0.43	0.71	-12.22	-233.83	209.39	0.61

	SWA p2	3.00	64.19	0.06	-0.78	-1.48	-0.08	0.03	-1.93	-90.60	86.75	0.82
	SWA p3	2.00	97.87	0.00	-0.81	-1.67	0.05	0.06				
	SWAM	3.00	91.43	0.00	0.12	-0.55	0.80	0.72	34.56	-115.76	184.88	0.21
	TT	1.00	0.00	1.00	0.57	-0.64	1.77	0.36				
	Between			0.00								
	Overall	35.00	94.96		-0.34	-0.71	0.04	0.08				
Ambulation and stairs												
		n	I-squared (between studies)	P-value	Effect size (Hedges' g) (95% CI)	Lower Limit	Upper limit	P-value				
	AGT3X	1.00	0.00	1.00	-0.35	-1.50	0.81	0.56				
	AW	3.00	78.96	0.01	0.00	-0.65	0.65	1.00	-10.13	-68.37	48.11	0.27
	BA	1.00	0.00	1.00	0.14	-1.02	1.31	0.81				
	BMC	2.00	0.00	0.36	0.05	-0.75	0.85	0.90				
	BP	1.00	0.00	1.00	1.15	-0.04	2.35	0.06				
	EPUL	1.00	0.00	1.00	-0.35	-1.50	0.80	0.55				
	FC	2.00	87.37	0.00	0.61	-0.23	1.45	0.15				
	FCHR	5.00	76.12	0.00	0.78	0.28	1.29	0.00	1.63	-11.54	14.79	0.72
	FF	3.00	82.83	0.00	1.10	0.43	1.77	0.00	5.87	-37.18	48.93	0.33
	GF225	1.00	0.00	1.00	0.87	-0.24	1.98	0.12				
	GF920XT	1.00	0.00	1.00	-0.50	-1.68	0.67	0.40				
	GVA	1.00	0.00	1.00	-0.16	-1.33	1.00	0.78				
	GVF	4.00	91.90	0.00	-1.24	-1.86	-0.62	0.00	-13.76	-19.72	-7.80	0.01
	GVS	1.00	0.00	1.00	-0.10	-1.26	1.07	0.87				

	JU	3.00	83.19	0.00	0.31	-0.34	0.95	0.35	-9.01	-165.97	147.94	0.60
	JU24	1.00	0.00	1.00	-0.09	-1.23	1.04	0.87				
	MB	1.00	0.00	1.00	-2.36	-3.67	-1.05	0.00				
	NF	1.00	0.00	1.00	-0.08	-1.21	1.04	0.88				
	PL	1.00	0.00	1.00	0.36	-0.81	1.53	0.54				
	Polar AW200	1.00	0.00	1.00	-0.69	-1.81	0.43	0.23				
	SWA	5.00	95.95	0.00	0.79	0.25	1.33	0.00	9.82	1.24	20.88	0.07
	SWA p2	2.00	96.06	0.00	0.18	-0.63	0.99	0.67				
	SWA p3	5.00	93.40	0.00	0.20	-0.32	0.71	0.46	6.93	-13.25	27.11	0.35
	SWAM	5.00	81.80	0.00	0.43	-0.09	0.94	0.10	-3.29	-18.40	11.81	0.54
	V	1.00	0.00	1.00	-1.42	-2.58	-0.27	0.02				
	WP	1.00	0.00	1.00	-2.72	-3.98	-1.46	0.00				
	WPO	1.00	0.00	1.00	-1.12	-2.44	0.20	0.10				
	Between			0.00								
	Overall	55.00	93.74		-0.09	-0.45	0.27	0.62				
Cycling												
		n	I-squared (between studies)	P-value	Effect size (Hedges' g) (95% CI)	Lower Limit	Upper limit	P-value				
	AW	1.00	0.00	1.00	-0.09	-1.54	1.35	0.90				
	AWS2	1.00	0.00	1.00	1.70	0.18	3.21	0.03				
	BMC	1.00	0.00	1.00	-0.53	-1.98	0.92	0.47				
	BP	1.00	0.00	1.00	-0.73	-2.21	0.74	0.33				
	FB	1.00	0.00	1.00	-0.91	-2.36	0.55	0.22				
	FC2	1.00	0.00	1.00	-1.05	-2.53	0.43	0.16				
	FCHR	2.00	97.72	0.00	-0.76	-1.83	0.31	0.16				

	GVF	1.00	0.00	1.00	-3.70	-5.55	-1.84	0.00				
	GVHR	1.00	0.00	1.00	0.15	-1.30	1.60	0.84				
	JU24	1.00	0.00	1.00	-3.67	-5.28	-2.05	0.00				
	MB	1.00	0.00	1.00	-0.90	-2.38	0.58	0.23				
	PA360	1.00	0.00	1.00	0.60	-0.86	2.06	0.42				
	SWA	3.00	89.39	0.00	-0.60	-1.45	0.25	0.17	-15.80	-409.77	378.16	0.70
	SWA p2	1.00	0.00	1.00	0.54	-0.92	1.99	0.47				
	SWA p3	3.00	54.95	0.11	-0.54	-1.38	0.31	0.21	-6.68	-54.80	41.43	0.32
	SWAM	1.00	0.00	1.00	-0.31	-1.75	1.14	0.68				
	TT	1.00	0.00	1.00	0.76	-0.70	2.23	0.31				
	WP	1.00	0.00	1.00	-3.89	-5.58	-2.19	0.00				
	Between			0.00								
	Overall	23.00	94.74		-0.73	-1.39	-0.06	0.03				
Running												
		n	I-squared (between studies)	P-value	Effect size (Hedges' g)	Lower Limit	Upper limit	P-value				
	AW	2.00	94.12	0.00	0.15	-0.70	1.00	0.73				
	BA	1.00	0.00	1.00	-0.61	-1.90	0.67	0.35				
	BMC	2.00	94.58	0.00	-0.15	-1.01	0.71	0.73				
	BP	1.00	0.00	1.00	0.20	-1.02	1.41	0.75				
	EPUL	1.00	0.00	1.00	0.07	-1.16	1.29	0.91				
	FC	2.00	88.84	0.00	-0.35	-1.23	0.54	0.45				
	FCHR	4.00	66.80	0.03	0.50	-0.11	1.11	0.11	-0.45	-22.54	21.64	0.94
	FF	2.00	58.63	0.12	0.60	-0.26	1.45	0.17				

	GF920XT	1.00	0.00	1.00	-0.27	-1.53	0.99	0.67				
	GVA	1.00	0.00	1.00	0.26	-1.01	1.53	0.69				
	GVF	3.00	46.39	0.15	-0.58	-1.33	0.17	0.13	-5.20	-171.72	161.34	0.76
	GVS	1.00	0.00	1.00	0.18	-1.08	1.44	0.78				
	JU	2.00	15.55	0.28	0.63	-0.24	1.50	0.16				
	JU24	1.00	0.00	1.00	0.92	-0.30	2.13	0.14				
	MB	1.00	0.00	1.00	0.17	-1.04	1.39	0.78				
	NF	1.00	0.00	1.00	0.63	-0.59	1.84	0.31				
	PL	1.00	0.00	1.00	0.21	-1.04	1.45	0.74				
	SWA	3.00	96.79	0.00	-0.14	-0.89	0.60	0.70	-0.73	-178.03	176.57	0.97
	SWA p2	1.00	0.00	1.00	-0.38	-1.59	0.83	0.54				
	SWA p3	2.00	88.85	0.00	-1.34	-2.22	-0.46	0.00				
	SWAM	2.00	94.20	0.00	0.10	-0.77	0.98	0.82				
	V	1.00	0.00	1.00	-0.23	-1.43	0.98	0.71				
	WP	1.00	0.00	1.00	-2.37	-3.68	-1.06	0.00				
	WPO	1.00	0.00	1.00	-0.42	-1.78	0.93	0.54				
	Between			0.04								
	Overall	38.00	92.05		-0.08	-0.41	0.25	0.65				
Sedentary and household												
		n	I-squared (between studies)	P-value	Effect size (Hedges' g) (95% CI)	Lower Limit	Upper limit	P-value				
	AW	3.00	97.07	0.00	0.29	-0.49	1.07	0.47	3.11	-389.99	396.22	0.93
	BMC	2.00	85.52	0.01	0.52	-0.41	1.45	0.27				

	BP	1.00	0.00	1.00	-0.78	-2.14	0.59	0.27				
	EPUL	1.00	0.00	1.00	-0.31	-1.65	1.02	0.65				
	FCHR	4.00	59.60	0.06	-0.14	-0.81	0.53	0.69	-0.31	-27.42	26.79	0.96
	FF	1.00	0.00	1.00	-0.30	-1.62	1.02	0.66				
	GF225	1.00	0.00	1.00	1.30	-0.04	2.64	0.06				
	GVF	1.00	0.00	1.00	-1.26	-2.64	0.12	0.07				
	JU	1.00	0.00	1.00	-0.85	-2.20	0.51	0.22				
	JU24	1.00	0.00	1.00	-0.94	-2.28	0.40	0.17				
	MB	1.00	0.00	1.00	-1.34	-2.75	0.08	0.06				
	SWA p2	2.00	0.00	0.60	0.71	-0.23	1.64	0.14				
	SWA p3	4.00	91.27	0.00	0.67	0.00	1.34	0.05	8.42	-16.91	33.74	0.29
	SWAM	5.00	97.42	0.00	0.04	-0.55	0.63	0.90	22.71	-42.47	87.89	0.35
	V	1.00	0.00	1.00	0.00	-1.32	1.32	1.00				
	WP	1.00	0.00	1.00	-0.62	-1.97	0.73	0.37				
	Between			0.06								
	Overall	30.00	94.84		-0.09	-0.51	0.32	0.66				
TEE (DLW)												
		n	I-squared (between studies)	P-value	Effect size (Hedges' g) (95% CI)	Lower Limit	Upper limit	P-value				
	EP	1.00	0.00	1.00	-0.68	-1.58	0.21	0.16				
	FF	1.00	0.00	1.00	-0.48	-1.35	0.38	0.27				
	GVF	1.00	0.00	1.00	-1.62	-2.63	-0.62	0.00				
	JU24	1.00	0.00	1.00	-1.96	-2.97	-0.95	0.00				
	MS	1.00	0.00	1.00	-0.68	-1.65	0.29	0.17				

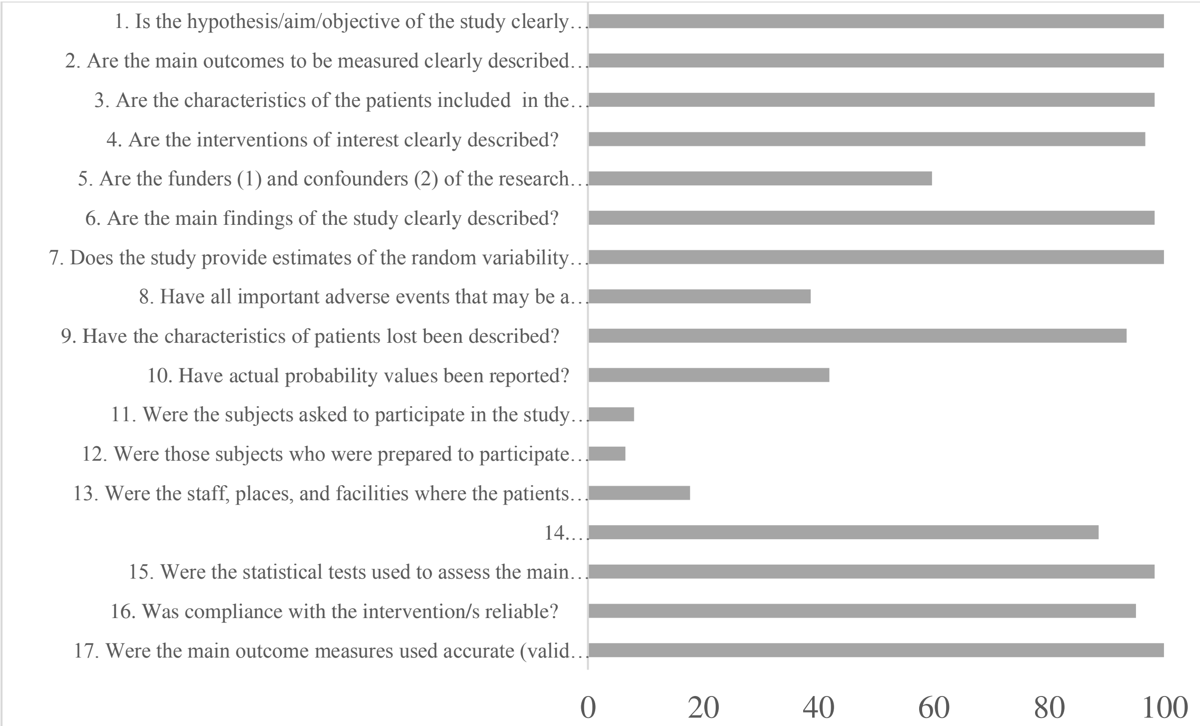
SWA	2.00	57.21	0.13	0.14	-0.46	0.75	0.65				
SWA p2	1.00	0.00	1.00	-0.23	-1.07	0.60	0.59				
SWA p3	5.00	94.20	0.00	-0.25	-0.64	0.13	0.19	7.03	-31.01	45.07	0.60
SWAM	2.00	0.00	0.99	-0.04	-0.64	0.56	0.90				
WPO	1.00	0.00	1.00	-1.68	-2.62	-0.75	0.00				
Between			0.00								
Overall	16.00	92.71		-0.68	-1.15	-0.21	0.00				

S8:

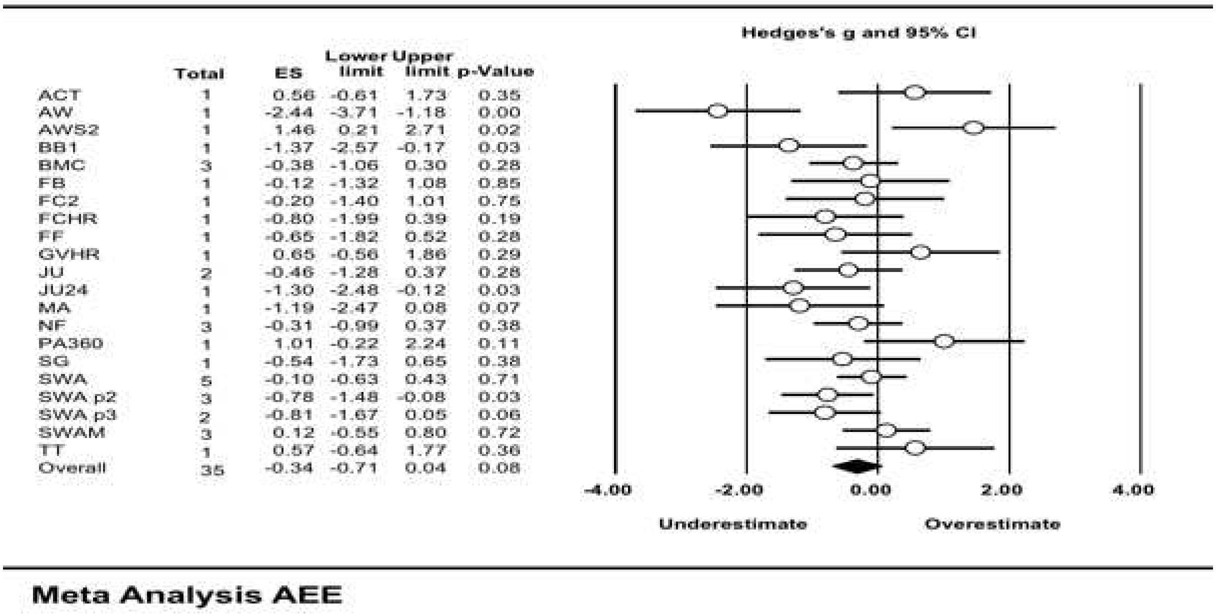
	Reporting (/11)	External validity (/3)	Internal validity (/4)
Alsubheen, 2016	10	0	4
Bai, 2017	9	0	4
Benito, 2012	8	0	4
Berntsen, 2010	9	0	4
Berntsen, 2012	9	2	4
Bhammar, 2016	11	0	4
Boudreaux, 2018	10	0	4
Brazeau, 2011	10	0	4
Brazeau, 2014	11	0	3
Brazeau, 2016	11	1	4
Brugniaux, 2010	8	1	3
Calabro, 2014	9	0	4
Calabro, 2015	11	1	4
Casiraghi, 2013	11	0	4
Choudhry, 2017	9	0	4
Colbert, 2011	10	1	3
Correa, 2016	10	0	3
Diaz, 2015	7	0	4
Diaz, 2016	9	0	4
Dondzilla, 2016	8	0	4
Dooley, 2017	10	0	4
Drenowatz, 2011	9	0	4
Erdogan, 2010	9	0	3
Fruin, 2010	9	0	3
Furlanetto, 2010	11	0	4

Gastin, 2017	8	0	4
Heiermann, 2011	8	2	4
Imboden, 2017	9	0	4
Jakicic, 2004	10	0	4
Johannsen, 2010	9	1	4
Kim, 2015	8	0	4
King, 2004	9	0	4
Koehler, 2011	10	1	4
Lee, 2011	9	0	4
Lee, 2014	9	0	4
Mackey, 2011	11	3	4
Martien, 2015	9	2	4
McMinn, 2013	9	0	4
Melanson, 2009	5	0	2
Mikulic, 2011	10	0	4
Montoye, 2017	10	0	4
Murakami, 2016	7	1	4
Nelson, 2016	10	0	4
Papazoglou, 2006	9	0	4
Price, 2017	9	0	4
Reece, 2015	9	0	4
Rousset, 2015	9	1	4
Ryan, 2013	10	0	2
Slinde, 2013	10	2	4
Smith, 2012	10	0	4
St-Onge, 2007	9	1	3
Stackpool, 2015	9	0	4
Tucker, 2015	11	0	4
Van helst, 2012	9	0	4
Van Hoye, 2014	9	0	4
Van Hoye, 2015	10	0	4
Vernillo, 2015	8	0	4
Wahl, 2017	9	0	4
Wallen 2016	9	0	4
Woodman, 2017	8	0	4

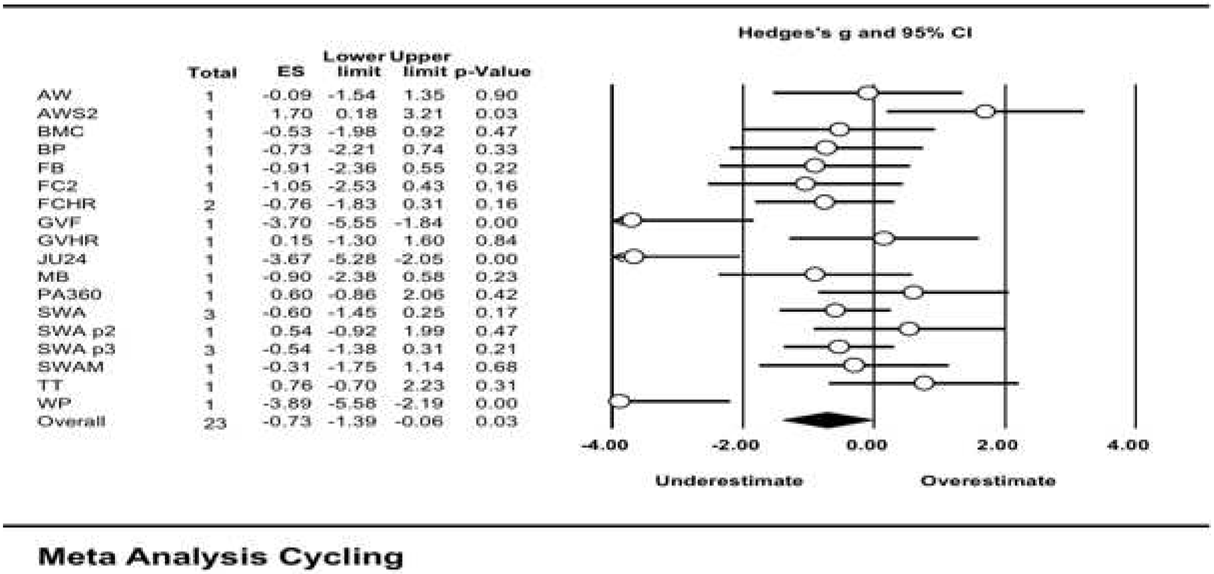
S9:



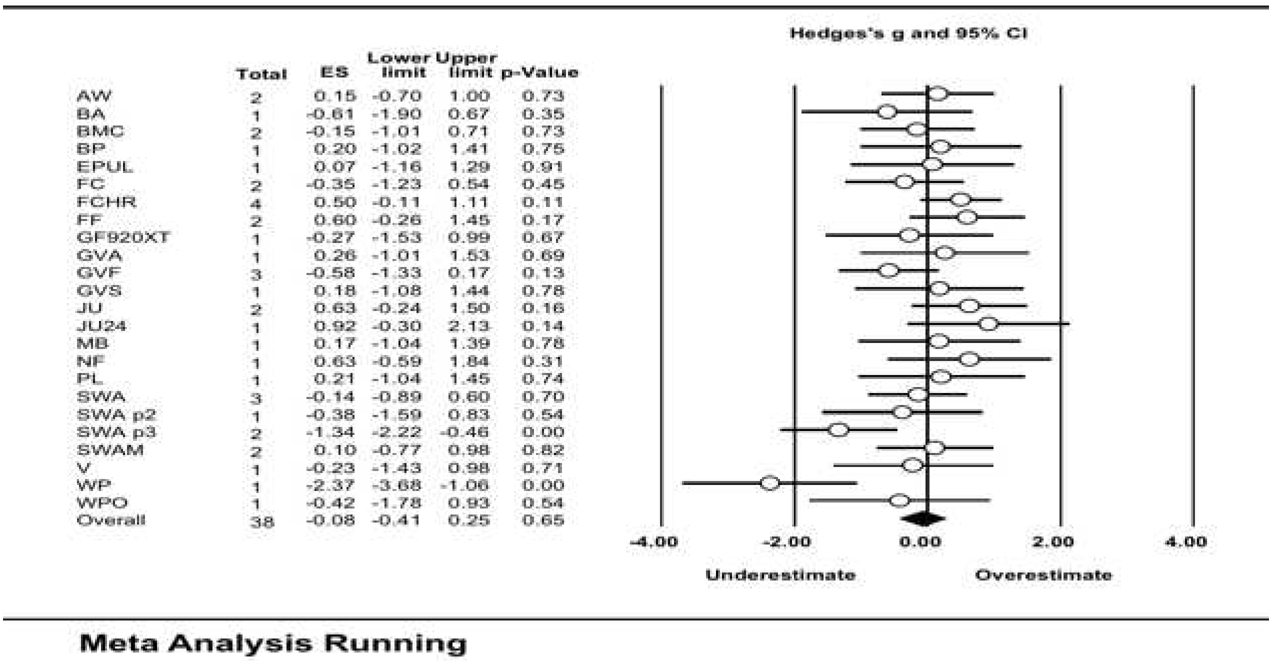
S10:



Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device for AEE. Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.



Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device during cycling. Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.



Pooled Hedges' g and 95% confidence intervals (CI) for estimates of energy expenditure relative to criterion measures per device during running. Total refers to number of effect sizes. A negative Hedges' g statistic represents an underestimation and a positive Hedges' g represents an overestimation.