Appendix 6a. Data extraction of studies with biological state outcomes.

No.	Author (year, location)	Study population (N; age ± SD or range [yrs]; % females; ethnicity)	Study duration	Wearable (placement; software; epoch-length; algorithm/cut-point)	Dimension (Outcome(s))	Criterion measures	Statistical analysis	Study conclusion	Overall risk of bias (low, some, high)	Funding/ Conflict of interest
7	Bélanger et al. (2013, Canada)	Toddler and preschool children (N=12; 3.1±1.0 yrs; 67% females; NR)	1 day	*Actiwatch-L* (non-dominant wrist, ankle; Actiwatch-L software; 30 sec; AlgoSmooth)	Biological state (Sleep/awake)	Polysomnography	Sensitivity; Specificity Accuracy; Negative predictive value	Despite the weak detection of wakefulness, Actiwatch-L appears to be a useful instrument for assessing sleep in preschoolers when used with an adapted algorithm.	Some	N/N
13	Camerota et al. (2018, United States)	Infants (N=90; 3.61 months; 43% females; African American (N=90))	1 day (1 night)	*Actiwatch-2* (left ankle; Phillips Actiware software v6.0; 15 sec; Actogram algorithm)	Biological state (Sleep time)	Observation (Video)	Correlations; Paired t-tests; Bland-Altman analysis; Cohen´s kappa; Sensitivity; Specificity	Actigraphy and sleep diaries were more concordant with one another than with videosomnography. Epoch-by-epoch analyses indicated that actigraphy had low sensitivity to detect wakefulness, compared with videosomnography.	Some	N/N
55	Quante et al. (2018, United States)	Healthy volunteers (N=13; 13.0±1.4 yrs; 46% females; White (n=2), Hispanic Black (n=2), Hispanic White (n=9))	5 days (1 night)	*Actigraph GT3X+* (non-dominant wrist; ActiLife 6; 1 minute; Cole–Kripke algorithm, Sadeh algorithm); *Actiwatch Spectrum* (non-dominant wrist; Respironics Actiware 5; 30 sec; medium threshold (40 counts per epoch))	Biological State (Total sleep time)	Polysomnography	Intraclass correlation coefficients, Accuracy, Sensitivity, Specificity	The two actigraphs provided comparable and accurate data compared to PSG, although both poorly identified wake episodes (i.e., had low specificity).	Some	N/N
69	Smith et al. (2020, New Zealand)	Children (N= 28; 7.2±1.2 yrs; 39% females; NZ Māori (7%), European (79%), Asian (14%))	1 day (1 night)	*Actical* (right hip, non-dominant wrist; Actical v3.0; NR; count-scaled algorithm, Sadeh algorithm); *ActiGraph GT3X+* (right hip, non-dominant wrist; Actilife v6.11.9; 15 sec; count-scaled algorithm, Sadeh algorithm)	Biological state (Total sleep time)	Polysomnography	Sensitivity; Specificity; Accuracy; Prevalence-adjusted bias-adjusted kappa; Bland-Altman analysis	Overall the count-scaled algorithm produced High sensitivity at the expense of low specificity in comparison with PSG. A best site placement for estimates of all sleep variables could not be determined, but overall the results suggested ActiGraph GT3X+ at the hip may be superior for sleep timing and quantity metrics, whereas the wrist may be superior for sleep quality metrics.	Some	NR/N
5	Barreira et al. (2015, United States)	Fourth-grade school children (N=34; 10.3±0.4 yrs; 55.8% females; NR)	7 days	*ActiGraph GT3X+* (waist; ActiLife software v6.5.2; 1 minute; Sadeh algorithm)	Biological state (Sleep time)	Sleep diary	Pearson correlation; Mean absolute percent differences; Paired t-tests; Bland-Altman analysis	The RSA is a refinement of our previous algorithm, allowing researchers who use a 24-h waist-worn accelerometry protocol to distinguish children's nocturnal	High	NR/NR

								sleep (including night time wake episodes) from daytime activities.		
9	Borghese et al. (2018, Canada)	Children (N=50; 10-13 yrs, 48% females; NR)	8 days	*Actical* (waist; NR; 15 sec; NR)	Biological state (Sleep efficiency)	Wearable (Actiwatch 2 (wrist))	Paired sample t-tests; Bland-Altman analysis; Pearson correlation	A waist-worn Actical accelerometer can accurately predict sleep efficiency in field settings among healthy 10- to 13-year-olds.	High	N[1]/N
29	Godino et al. (2020, United States)	Children (N=26; 9.9±0.7 yrs; 50% females, NR)	1 day	*Fitbit Charge HR* (non-dominant wrist; Fitabase; 1 minute; proprietary algorithms)	Biological state (Sleep/wake pattern)	Polysomnography	Mean absolute percentage error; Accuracy; Cohen's kappa; Sensitivity; Specificity; Bland-Altman analysis	Fitbit Charge HR had adequate sensitivity in classifying sleep, but had limitations in detecting wake, and was more accurate in detecting heart rate during sleep than during exercise, in healthy children.	High	N/NR
41	Kinder et al. (2012, United States)	Fith graders (N=22; 10-11 yrs; 32% females; NR)	3 days	*ActiGraph GT1M* (left hip; Actilife Lifestyle Monitor v3.6.0; 1 minute; NR)	Biological state (Total sleep time)	Wearable (Mini-MotionLogger (wrist))	Pearson correlation; Paired t-test,	A hip-worn accelerometer may provide a cost-effective means of gathering physical activity and sleep data simultaneously in large samples of children with or without an accompanying sleep diary.	High	N/NR[3]
46	Meltzer & Westin (2011, United States)	School children (N=38; 9.92±1.5 yrs; 55% females; Caucasian (N=29), Afrcian-American (N=7), Hispanic (N=1))	7 days	*Motionlogger Sleepwatch* (non-dominant wrist; Action-W software; 1 minute; Sadeh algorithm)	Biological state (Sleep time)	Sleep Diary	Repeated measures ANOVA; Pearson correlations	Established standards for the scoring and reporting of actigraphy data are needed. Until these are in place, researchers need to be conscientious when using actigraphy and clearly report scoring rules and variables.	High	NR/N
53	Pesonen & Kuula (2018, Finland)	Children and adolescent (N=34; younger group: 11±0.8 yrs; 53% females; older group: 17.8±1.8 yrs; 47% females; NR)	1 day	*Polar A370 prototype* (non-dominant wrist; Polar Software; 30 sec; proprietary algorithm); *Actiwatch 2* (non-dominant wrist; Actiware 6.0.8; 30 sec; default medium sensitivity threshold (40 counts per epoch))	Biological state (Sleep time)	Polysomnography	Paired t-test; Bland-Altman analysis	Consumer-targeted wrist-worn device performs as well as, or even better than, the previously validated AW2 against PSG in children and adolescents.	High	Y[2]/N
68	Sitnick et al. (2008, United States)	Children with autism & healthy (N=58; 47.8±12.7 months; 50% females; Caucasian (60%), Hispanic (10%), African American (4%), Asian American (4%), and mixed (22%))	2 days (2 nights)	*Actiwatch 64* (non-dominant ankle; NR; 1 minute; smoothing algorithm)	Biological state (Total sleep time)	Observation (Video)	Agreement; Sensitivity; Specificity; Paired t-test; Spearman correlations	Actigraphy has poor agreement for detecting nocturnal awakenings, compared with video observations, in preschool-aged children.	High	N/N
72	Unno et al. (2021, Japan)	Newborns (N=40; 35.3–45 weeks; 50% females; NR)	1 day (1 night)	*Actiwatch 2* (ankle; NR; 1 minute; NR)	Biological state (Sleep-wake states)	Polysomnography	Mixed-effects logistic regression; Receiver operating characteristic (ROC)	Study suggests that sleep–wake states of NICU-hospitalised newborns can be precisely	High	NR/N

							curve analysis Sensitivity; Specificity; Positive predictive value; Negative predictive value	determined using actigraphy on the ankle.		
[1]N= No; [2]NR=Not reported; [3]Y=Yes										

Appendix 6b. Data extraction of studies with posture/activity type outcomes.

No.	Author (year, location)	Study population (N; age ± SD or range [yrs]; % females; ethnicity)	Study duration	Wearable (placement; software; epoch-length; algorithm/cut-point)	Dimension (Outcome(s))	Criterion measures	Statistical analysis	Study conclusion	Overall risk of bias (low, some, high)	Funding/ Conflict of interest
1	Alghaeed et al. (2013, United Kingdom)	Preschool children (N=32; 4.1±0.5 yrs; 67% females, NR)	< 1 day (1 hr)	*ActivPAL* (right tHigh; activPAL Professional Research Edition v5.8.2.3; 15 sec; NR)	Posture/Activity Type (sitting; sedentary breaks)	Observation (Video)	Bland-Altman analysis; Paired t-test	A default setting of 10 s for the activPALTM appears unsuitable for quantification of breaks in sitting in young children, in whom a minimum sitting/upright period of 2 s will provide much Higher accuracy with minimal bias.	High	N[1]/N
10	Busser et al. (1997, Germany)	Children (N=10; 4-10 yrs; 20% females; NR)	< 1 day (mean 55 minutes)	*Dynaport ADL Monitor* (backpack; ADL monitor software; NR; NR)	Posture/Activity type (standing, sitting, lying, locomotion, swing)	Observation (Video)	Sensitivity; Predictive value	The overall minimal and maximal validity are 73.15±1.96 x 4.48 and 91.31±1.96 x 1.75 weighed standard deviation.	High	NR[3]/NR
14	Carlson et al. (2019, Australia)	Children (N=195; 10.5±0.7 yrs; 48.7 % females; NR)	8 days	*ActiGraph GT3X+* (hip; ActiLife Software 2012; 15 sec and 1 minute; ≤25, ≤75, ≤100, ≤300 cpm)	Posture/Activity Type (Sedentary pattern)	Wearable (ActivPAL (tHigh))	Mean differences; Mean absolute deviation; Intraclass correlation coefficients	Sedentary patterns derived from two commonly used ActiGraph cut-points did not appear to reflect postural changes.	High	N/N
20	Davies et al. (2012, United Kingdom)	Children (N=20; 4.1 yrs; 66% females; NR)	< 1 day (1 hr)	*ActivPAL* (tHigh; activPAL Professional Research Edition software v 5.8.2.3; NR; NR)	Posture/Activity type (sitting, lying, standing, walking)	Observation (Video)	General Linear Model ANOVA; Bland-Altman analysis	The activPAL had acceptable validity, practical utility, and reliability for the measurement of posture and activity during freeliving activities in pre-school children.	High	N/Y[2]
22	De Decker et al. (2013, Belgium)	Preschool children (N=44; 5.49±0.59 yrs; 50% females; NR)	< 1 day (1 hr)	*ActivPAL* (tHigh; ActivPAL software v6.0.8; 15 sec; NR); *ActiGraph GT1M* (right hip; Actilife software v 5.4.6 Lite Edition; 15 sec; NR)	Posture/Activity Type (Sedentary time; non-sedentary time)	Observation (Video)	Two-way repeated-measures ANOVA; Bland-Altman analysis	Low classification accuracy was found for the ActivPAL and the GT1M ActiGraph to measure sedentary behavior in preschoolers. No correction factor can be suggested to make the sedentary estimates of the GT1M ActiGraph and the ActivPALi convergent as no systematic bias and wide limits of agreement were found.	High	N/N
36	Hurter et al. (2018, United Kingdom)	Children (N=21; 10.2±0.3 yrs; 61.9% females; NR)	2 days	*GENEActiv* (each wrist; GENEActiv software v3.1; NR; ENMO metric); *ActiGraph GT9X* (each wrist; ActiLife v6.13.3; NR; ENMO metric); *ActiGraph GT3X* (right	Posture/Activity type (Time spent sedentary or stationary)	Wearable (activPAL (tHigh))	Paired t-tests; Cohen´s d; Bland-Altman analysis; Pearson correlations, mean percent errors; Mean absolute	The stationary thresholds underestimated stationary time when applied to free-living data in relation to activPAL.	High	N/N

				hip; ActiLife v6.13.3; NR; ENMO metric)			percent errors; Equivalence testing			
37	Hurter et al. (2019, United Kingdom)	Children (N=21; 10.2±0.3 yrs; 61.9% females; NR)	2 days	*GENEActiv* (non-dominant wrist; GENEActiv software v3.1; 15 sec; Sedentary Sphere posture algorithm); *ActiGraph GT9X* (non-dominant wrist; ActiLife v6.13.3; 15 sec; Sedentary Sphere posture algorithm)	Posture/Activity type (Sedentary Sphere)	Wearable (activPAL (tHigh))	Agreement; Sensitivity; Specificity; Bland-Altman analysis; Equivalence test; Mean percent errors; Mean absolute percent errors	Sedentary Sphere can be used to classify posture from wrist-worn AG and GA accelerometers for group-level estimates in children, but future work is needed to improve the algorithm for better individual-level results.	High	NR/N
45	Kwon et al. (2019, United States)	Toddlers (N=24; 13-35 months; 50% females; NR)	< 1 day (between 8 and 25 min)	*ActiGraph wGT3X-BT* (hip, wrist; ActiLife software v6.13; 5 sec; Fast Fourier transform algorithm)	Posture/Activity type (run, walk, crawl, climb up & down, stand, sit)	Observation (Video)	Mann-Whitney tests; Sensitivity index d	Hip vertical axis counts alone may be unable to capture walking as physical activity and "carried" as sedentary behavior among toddlers. Machine learning techniques that utilize additional accelerometry signal features could help to recognize behavior types, especially to differentiate being "carried" from ambulatory movements.	High	N/N
49	Nam & Wook Park (2013, South Korea)	Baby and child (N=3;16-20 months; NR; NR)	< 1 day	*SCA3000* (waist; NR; NR; Naive Bayes, Bayes Net, Support Vector Machine, k-Nearest Neighbor, Decision Tree, Decision Table, Multilayer Perceptron, Logistic: Logistic Regression)	Posture/Activity type (e.g., wiggling, standing still, sitting down, crawling)	Observation (Video)	Accuracy of Naïve Bayes, Bayes Net, Support Vector Machine, k-Nearest Neighbor, Decision Tree, Decision Table, Multilayer Perceptron, Logistic classifiers	The overall accuracy of activity recognition was 96.2% using only a single wearable triaxial accelerometer sensor with the k-Nearest Neighbor.	High	N/NR
50	Narayanan et al. (2020, New Zealand)	Children (N=15; 10.0±2.6 yrs; 33.3% females; NR)	< 1 day (2 hrs)	*Axivity AX3* (lower back, wrist, tHigh; OmGui v1.0.0.30; 5 sec; random forest)	Posture/Activity type (sedentary activities, ambulatory activities, dynamic standing)	Observation (Video)	Leave-one-out cross-validation; Sensitvity; Specificity; Accuracy	This validation study demonstrated that a dual-accelerometer system previously validated in a laboratory setting also performs well in semi free-living conditions.	High	NR/N
61	Ruch et al. (2011, Switzerland)	Healthy children: training data set group (N=21; 10.8±1.3 yrs; 57% females; NR); testing data set (N=20; 10.6±1.6 yrs; 60% females; NR)	7 days	*ActiGraph GT1M* (wrist, hip; NR; 1 sec; k-nearest neighbor, normal density discriminant function custom decision tree)	Posture/Activity type (e.g., stationary activities, walking, running)	Observation (Video)	Classification process combined three different classifiers such as k-nearest neighbor (k-NN), normal density discriminant function (NDDf) and a custom decision tree (CDT)	The presented classification procedure provides additional information on the PA behavior in children registered by established accelerometers.	High	N/N
65	Silva et al. (2019, Portugal)	Children (N=49; NR; NR; NR)	7 days	*ActiGraph GT3X+* (right hip; ActiLife v6.10.4; 15 sec; < 100 cpm)	Posture/Activity type (Time spent sitting/lying)	Wearable (ActivPAL (tHigh))	ANOVA; Paired t-tests; Wilcoxon; Intraclass correlation	GT3X (cut-point of <100 counts·min−1) overestimated sedentary time of free-living	High	N/NR

							coefficient; Regression; Bland-Altman analysis	activities and did not detect changes resulting from a classroom standing desk intervention in adolescents.		
[1]N= No; [2]NR=Not reported; [3]Y=Yes										

Appendix 6c. Data extraction of studies with intensity outcomes.

No.	Author (year, location)	Study population (N; age ± SD or range [yrs]; % females; ethnicity)	Study duration	Wearable (placement; software; epoch-length; algorithm/cut-point)	Dimension (Outcome(s))	Criterion measures	Statistical analysis	Study conclusion	Overall risk of bias (low, some, high)	Funding/ Conflict of interest
12	Calabro et al. (2013, United States)	Healthy youth (N=28; 12.4±1.6 yrs; 46% females; 76% Caucasian (with 14% Hispanic and 10% Asian))	14 days	*SenseWear Pro3* (upper arm; Research Software 6.1; NR; algorithms 2.2), *SenseWear Mini* (left arm; Research Software 7.0; NR; algorithms 2.2)	Intensity (Energy expenditure)	Doubly labeled water	Mixed-model ANOVA; Pearson correlation; Bland-Altman analysis	The newly developed SenseWear Armband 5.0 algorithms outperformed the version 2.2 algorithms for group comparisons.	Low	Y[2]/N[1]
31	Hallal et al. (2013, Brazil)	Adolescents (N=25; 13±0.3 yrs; 64% females; NR)	4 days	*ActiGraph GT1M* (waist; NR; 5 sec; MVPA ≥ 2296 cpm)	Intensity (Total energy expenditure; Physical activity energy expenditure)	Doubly labeled water	Spearman coefficients; Linear regression models	Objectively measured physical activity significantly contributes to the explained variance in both TEE and PAEE in Brazilian youth.	Low	N/N
38	Ishikaw-Takata et al. (2013, Japan)	Participants from junior High school (N=60; 12-15 yrs; 47% females; NR)	7 days	*Actimaker EW4800P* (waist; commercially available accelerometry software; NR; Equation reported)	Intensity (Total energy expenditure; Steps)	Doubly labeled water	Paired t test; Spearman correlations; Intraclass correlation coefficient; Bland-Altman analysis	While accelerometry estimated TEE accurately, it did not provide the precise measurement of PAEE and PAL.	Low	N/N
44	Krishnaveni et al. (2009, India)	Indian children (N=58; 8.7±0.3 yrs; 48% females; NR)	14 days	*ActiGraph AM7164, ActiGraph GT1M* (right hip; software programme MRC Epidemiology Unit; 1 minute; Equation reported)	Intensity (Total energy expenditure)	Doubly labeled water	Pearson correlation; Linear regression; Kappa statistics; Bland-Altman analysis	Activity measured using Actigraph accelerometers was not related to TEE and PAL derived using the DLW technique in children in Mysore. Actigraphs may not be useful in predicting EE in this setting, but may be better used for judging activity patterns.	Low	N/N
51	Ojiambo et al. (2012, United Kingdom, Belgium, Sweden, Spain)	Children (N=49; 6.9±1.5 yrs; 51% females; NR)	7 days	*ActiTrainer* (right hip, NR; 15 sec; Evenson cut points); *3DNX model v3* (right hip; NR; 15 sec; Evenson cut points)	Intensity (Total energy expenditure)	Doubly labeled water	Hierarchically nested regression models; Ordinary least squares regression; Prediction errors (leave-one-out cross validation)	The comparative validity of hip-mounted uniaxial and triaxial accelerometers for assessing PA and EE is similar.	Low	N/N
57	Reilly et al. (2006, United Kingdom)	Healthy young children (N=85; 4.6±1.1 yrs; 40% females; NR)	7-10 days	*ActiGraph CSA/MT1* (right hip; NR; 1 minute; Ekelund equation, Puyau equation)	Intensity (Total energy expenditure)	Doubly labeled water	Bland-Altman analysis	Simple approaches using the Actigraph appear to be inadequate for the estimation of free-living TEE in young children at present.	Low	N/NR[3]
64	Sijtsma et al. (2013, Netherlands)	Preschool children (N=30; 3.5±0.3 yrs; 60% females)	5 days	*Tracmor$_D$* (lower back; NR; NR; equation reported)	Intensity (Total energy expenditure)	Doubly labeled water	Pearson correlations; Linear regression	Tracmor$_D$ provides moderate-to-strong validity evidence that	Low	Y/N

							analyses; Bland-Altman analysis	supports its use to evaluate PAL and AEE in preschool children.		
4	Bäcklund et al. (2010, Sweden)	Overweight, and obese children (N=22; 10.3±0.99 yrs; 50% females; NR)	14 days	*SenseWear Pro2* (right arm; InnerView Professional; NR; proprietary algorithms)	Intensity (Total energy expenditure)	Doubly labeled water	Paired t-test; Bland-Altman analysis; Pearson correlations	The SWA together with software version 5.1, but not 6.1, is a valid method for accurately measuring energy expenditure at group level of free-living overweight and obese children.	Some	N/N
15	Carter et al. (2008, United Kingdom)	Adolescents (N=23; 15±1 yrs; 26.1% females, NR)	10 days	*3dNX accelerometer model v2* (right hip; NR; 1 minute; Log-linear TEE model)	Intensity (Energy expenditure)	Doubly labeled water	Bland-Altman analysis	The 3dNXTM accelerometer can be used to predict free-living daily energy expenditure with a standard error of estimate of 1.65 MJ in adolescents.	Some	N/NR
17	Costello et al. (2019, United Kingdom)	Professional young male rugby players (N=8; 17±1 yrs; 0% females; NR)	14 days (pree-season); 7 day (in-season)	*SenseWear Pro 3* (upper arm; Internal proprietary algorithm v5.2; 1 minute; proprietary algorithm)	Intensity (Energy expenditure)	Doubly labeled water	Mean bias; Typical error of the estimate; Pearson correlation; Paired t-tests	Findings demonstrate the limitations of utilising isolated or combined wearable technology to accurately determine the total energy expenditure of professional collision-based sport athletes across different stages of the season.	Some	NR/NR
35	Hoos et al. (2003, Netherlands)	Healthy children (N=11; 6.9±2.2 yrs; 73% females; NR)	14 days	*Tracmor2* (waist; NR; 1 minute; NR)	Intensity (Counts)	Doubly labeled water	Linear regression analysis	Tracmor2 is a valid instrument to measure physical activity in children under free-living conditions.	Some	NR/NR
47	Migueles et al. (2019, Sweden)	Preschool children (N=39; 5.5±0.1 yrs, 46.2% females; NR)	7 days	*ActiGraph wGT3X-BT* (non-dominant wrist; ActiLife 6.2.0 software; 5 sec; ENMO, BFEN, HFEN, HFEN+, MAD)	Intensity (Activity energy expenditure; Total energy expenditure)	Doubly labeled water	Independent t-tests; Bias and root mean square error (rMSE); Bland-Altman analysis	Euclidian norm minus 1g and the other alternate summary metrics explained more of the variance in TEE and AEE than the ActiGraphs activity counts in 5-yr-old children.	Some	N/N
58	Rodriquez et al. (2002, France)	Non-obese children and adolescents (N=20; 11.15±3.8 yrs; 35% females; NR)	1 day	*Tritrac-R3D* (waist; Tritrac-R3D software; NR; proprietary formula)	Intensity (Energy expenditure)	Indirect room calorimetry	Paired t tests; Linear regression analysis; Bland-Altman analysis; Spearman correlations	Correlation between the vector magnitude generated by the TriTrac-R3D accelerometer and EE of activities derived from HR monitoring is High. When compared with the HR method, the TriTrac-R3D and activity diary are not systematically accurate and must be carefully used for the assessment of children's EE depending on the purpose of each study.	Some	N/NR
62	Santos et al. (2014, Portugal)	Basketball player from Junior National Team (N=12; 16.4±0.5 yrs; 66.6% females; NR)	7 days	*Actiheart* (chest; commercial software v.4.0.46; 1 minute; NR)	Intensity (Energy expenditure)	Doubly labeled water	Concordance coefficient correlation; Pearson correlation; Bland-Altman analysis	ACC + HR models are a valid alternative to estimate TEE but not AEE in a group of Highlyactive individuals however the considerable rate of equipment failure (~50%) limits its usefulness.	Some	N/NR
2	Alhassan et al. (2017,	Preschool children (N=74; 4.4±0.8 yrs; 36% females; NR)	< 1 day (30 minutes)	*Actiwatch spectrum* (non-dominant wrist; NR; 15 sec; sedentary ≤79, light	Intensity (Time spent in different intensities	Observation (Direct)	Spearman correlation; Agreement; Sensitivity;	The AW-E and AG-P estimated times spent in MVPA were similar to DO, but the weak agreement	High	N/N

	United States)			80–261, moderate 262–405, and vigorous ≥406); *ActiGraph GT3X* (lower back; NR; 15 sec; sedentary ≤301, light 302–614; moderate 615–1230, and vigorous ≥1231)	(sedentary, LPA, MVPA))		Specificity; Kappa; Positive and negative predictive value; Bland-Altman analysis	statistics indicate that neither device cut-point equations provided accurate estimates at the individual level.		
3	Allor & Pivarnik (2000, United States)	Sixth-grade girls (N=46; 12±0.6 yrs; 100% females; NR)	4 days	*Caltrac accelerometer* (hip; NR; NR; NR)	Intensity (Energy expenditure)	Wearable (HR monitor (wrist))	Pearson correlation; Bland-Altman analysis	HR and PAR were stable across 2 d. PAR underestimated caloric expenditure by approximately 14%. Caltrac showed the least utility in both reliability and validity.	High	NR/NR
6	Beets et al. (2011, United States)	Children (N=149; 8.6±1.7; 40% females, NR)	< 1 day (5.7±0.8 hrs)	*Walk4Life* (right hip; NR; 5 sec; 120 steps·min^{-1})	Intensity (Time spent in MVPA)	Wearable (ActiGraph GT1M (waist))	Bland-Altman analysis; Pitman's test of differences; Paired t-test	Pedometer-determined MVPA provided comparable estimates of MVPA for older children (10–14-year-olds).	High	NR/NR
8	Bock et al. (2010, Germany)	Preschool children (N=33; males: 4.3±1.1 yrs, females: 3.5±0.8 yrs; 36% females; NR)	< 1 day (approx. 1-2 hrs)	*Actiheart* (chest; NR; 15 sec; NR)	Intensity (Time spent sedentary and MVPA)	Observation (Direct)	Accuracy; Sensitivity; Specificity	Devices that combine HR and ACC data yield an accurate classification of MVPA in preschoolers but perform less well for classifying SB.	High	N/N
11	Byun et al. (2018, United States)	Preschool children (N=27; 4.9±1.0 yrs; 41% females; NR)	1 day	*Fitbit Flex* (non-dominant wrist; Fitabase; 1 minute; proprietary algorithm)	Intensity (Time spent in sedentary behavior, MVPA, and total PA)	Wearable (ActiGraph GT3X+ (hip))	Pearson correlation; Mean absolute percent errors; Repeated measures ANOVA	The FF activity monitor accurately estimated the amount of time spent in SED and overall PA in preschool-aged children, but with an underestimation of moderate-to-vigorous PA.	High	NR/N
16	Coe & Pivarnik (2001, United States)	High school male basketball players (N=10; 12.8±0.4 yrs; 0% females; NR)	< 1 day (approx. 55 min)	*CSA Accelerometer* (right hip; NR; 1 minute; NR)	Intensity (Counts)	Observation (Direct)	Pearson correlation; repeated measure ANOVA	The CSA provides valid estimates of PA intensity (compared to CARS and HR) during basketball played by adolescent boys.	High	NR/NR
18	Crouter et al. (2015, United States)	Children (N=42; 12.6±0.8 yrs; 35.7% females; Hispanic 21.4%, Black/African 64.2%, Asian 19.2%, White 16.6%)	< 1 day (Approx. 2 hrs)	*ActiGraph GT3X+* (dominant wrist; ActiLife Software; 1 sec; Wrist VA regression model, wrist VM regression model)	Intensity (Energy expenditure (MET)); Time spent in different intensities (sedentary, LPA, MVPA))	Indirect calorimetry	One-way repeated-measures ANOVA; Root mean squared error; Mean bias	Compared to measured values, the VA and VM regression models developed on wrist accelerometer data had insignificant mean bias for child-METs and time spent in SB, LPA, MPA, and VPA; however, they had large individual errors.	High	N/N
19	Crouter et al. (2018, United States)	Children (N=42; 12.6±0.8 yrs; 35.7% females; Hispanic 21.4%, Black/African 64.2%, Asian 19.2%, White 16.6%)	< 1 day (Approx. 2 hrs)	*ActiGraph GT3X+* (ankle; ActiLife Software; 1 sec; VM2RM)	Intensity (Energy expenditure (MET); Time spent in different intensities (sedentary, LPA, MVPA))	Indirect calorimetry	Paired t-tests; Root mean squared error; Mean bias	Compared to the K4b2, the ankle VM2RM provided similar estimates to measured values during unstructured play and provides a feasible wear location for future studies.	High	N/Y

21	De Craemer et al. (2015, Belgium)	Preschoolers (N=41; 5.43±0.63 yrs; 48.8% females; NR)	4 days	*Omron Walking StylePro HJ-720IT-E2* (hip; Omron Health Management Software version E1.012; NR; NR)	Intensity (Steps)	Wearable (ActiGraph GT1M (hip))	Pearson correlations; Intraclass correlation coefficients; Independent sample t-test; Paired sample t-test; Bland-Altman analysis	Both the accelerometer-based as pedometer-based step counts are valid estimates of preschoolers' physical activity levels during free-living activities based on group estimates. High agreement between both step counts justifies combining and comparing pedometer- and accelerometer-based step counts.	High	N/NR
23	Djafarian et al. (2013, United Kingdom)	Children (N=42; 4.06±0.73 yrs; 47.62% females; Caucasian (N=42))	< 1 day (2 hrs)	*Actiwatch-L* (non-dominant wrist; Actiwatch analysis software; 1 minute; NR)	Intensity (Counts)	Observation (Direct)	Correlation coefficient; Linear regression	These data suggest that the Acti-watch (a wrist worn accelerometer) is a valid tool for assessing levels of physical activity in young chil- dren.	High	N/N
24	Duncan et al. (2011, New Zealand)	Children (N=114; 8.4±1.8 yrs; 55% females; NR)	< 1 day (5 hrs)	*New Lifestyle 1000* (hip; NR; 4 sec; NR)	Intensity (Steps; Time spent in MVPA)	Wearable (Actical (hip))	T-test; Bland-Altman analysis	Compared with a validated omnidirectional accelerometer, however, the NL-1000 significantly underestimates the total minutes of MVPA accumulated by 5–11-yearold children during a normal school day.	High	NR/NR
25	Etienne et al. (2016, United States)	Multiethnic preschoolers (N=30; 3.5±0.6 yrs; 43% females; 46% native Hawaiian, 14% other Pacific Islander, rest not reported)	< 1 day (Approx. 12 hrs)	*Actical accelerometer* (non-dominat wrist; NR; 15 sec; Sedentay (≤40 cpm), light (≥41 or ≤2295 cpm; moderate ≥2296 or ≤6815 cpm; vigorous ≥6816 cpm)	Intensity (Counts; Time spent in different intensities (sedentary, LPA, MVPA))	Observation (Direct)	Intraclass correlation coefficients; Kappa statistics	Accelerometers can be objective, valid, and accurate physical activity assessment tools compared to conventional PA logs and subjective reports of activity for preschool children of mixed ethnicity.	High	N/N
26	Finn & Specker (2000, United States)	Children (N=40; 3-4 yrs; 60% females; Caucasian (N=38))	< 1 day (5-6 hrs)	*Actiwatch activity monitor* (waist; NR; NR; NR)	Intensity (Time spent in different intensities (sedentary, LPA, MVPA))	Observation (Direct)	Correlation; Mixed model repeated measures analysis	3-min CARS score correlates with 3-min activity counts, favoring the use of the activity monitors in assessing physical activity in preschool-aged children.	High	N/NR
27	Gao et al. (2010, United States)	School children (N=225; 12.48 ±1.01 yrs; 50% females, NR)	< 1 day (1 ½ hrs)	*Yamax Digi-Walker SW-701* (left hip; NR; NR; NR)	Intensity (Time spent in MVPA)	Wearable (Actical (hip))	Pearsons correlation; Kappa agreement	Pedometer-based steps per minute is a valid tool to survey physical activity time against accelerometers.	High	NR/NR
28	Garcia-Prieto et al. (2017, Spain)	Children (N=32; 9.9±0.6 yrs; 62.5% females; Caucasian (N=32))	< 1 day (1 ½ hrs)	*ActiGraph GT1M* (right hip; Actilife v6.01; 5 sec; NR)	Intensity (Energy expenditure; Counts)	Indirect calorimetry	Correlation coefficients; Student's t-test; chi-square test	Accelerometer and HR monitors are useful devices for estimating EE during endurance games, but only HR monitors estimates are accurate for endurance games.	High	N/N
30	Grydeland et al. (2014, Norway)	Children (N=18; 9.9±0.3 yrs; 66.6% females; NR)	3 days	*ActiGraph GT3X+* (waist; ActiLife software v5.5.5; 10 sec; sedentary <100 cpm, LPA 100 < 2000 cpm, moderate 2000 < 6000, vigorous ≥6000 cpm); *ActiGraph AM7164*	Intensity (Time spent in different intensities (sedentary, LPA, MVPA))	Wearable (ActiGraph GT1M (waist))	Two-way mixed-model ANOVA; Intraclass correlation coefficient; Bland-Altman analysis	The ActiGraph model AM7164 yields Higher outputs of mean physical activity intensity (mcpm) than the models GT1M and GT3X+ in children in free-living conditions. The generations GT1M and GT3X+ provided comparable outputs.	High	NR/N

				(waist; DOS-based program (RUI24, v2.13B); 10 sec; sedentary <100 cpm, LPA 100 < 2000 cpm, moderate 2000 < 6000, vigorous ≥6000 cpm)						
32	Hands et al. (2006, Australia)	Healthy children (N=24; 66.6±3.5 months; 50% females; NR)	5 days (per day 30 min)	*ActiGraph AM7164* (waist; NR; 10 sec; NR); *Yamax Digi-Walker SW-200* (waist; NR; NR; NR)	Intensity (Energy expenditure; Time spent physically active)	Observation (Direct)	Pearson correlations; T-tests; One-way ANOVA	When the children were grouped into low, moderate, and High activity levels using observation, the pedometer data were better able to separate the groups than the accelerometer data. These findings indicate that the pedometer is a better measure of free play physical activity in 5- and 6-year-old children compared to the accelerometer.	High	NR/NR
33	Hart et al. (2011, United States)	Children (N=36; 10.2±3.1 yrs; 58.3% females; Caucasian (70%), Ethnic minority (including African American, Hispanic, Native American (30%))	< 1 day (Approx. 8 hrs)	*New Lifestyle NL-1000* (waist; NR; 4 sec; Default setting (MVPA 4-9); *Omron HJ-151* (waist; NR; NR; roprietary algorithm); *Walk4Life MVP* (waist; NR; NR; 100 steps/minute); *Yamax Digi-Walker SW-200* (waist; NR; NR; NR)	Intensity (Time spent in MVPA; Steps)	Wearable (AcgtiGraph GT1M (waist))	Mean absolute percent error; Repeated-measure ANOVA; Bland-Altman analysis; Spearman correlations	Low-cost instruments may be useful for measurement of both MVPA and Steps in children´s physical activity interventions and program evaluation.	High	NR/NR
34	Hislop et al. (2012, United Kingdom)	Preschool children (N=31; 4.4±0.8 yrs; 51.6% females; NR)	< 1 day (1 hour)	*ActiGraph GT1M* (waist; NR; 15 sec; >615 (3 yrs), >812 (4 yrs) cpm); *RT3* (waist; NR; 15 sec; walking relaxed > 413 cpm; light jog > 780 cpm)	Intensity (Time spent in MVPA)	Observation (Direct)	Pearson correlations; Spearman correlations; Freidman's repeated Measures ANOVA; Wilcoxon paired t-test; Bland-Altman analysis	There was no advantage of a triaxial accelerometer over a uniaxial model. Shorter epochs result in significantly Higher number of minutes of MVPA with smaller bias relative to direct observation.	High	N/NR
39	Janz (1994, United States)	Children (N=31; 11.2±2 yrs; 48% females; NR)	3 days	*CSA accelerometer* (waist; NR; 1 minute; NR)	Intensity (Counts)	Heart telemetry	Pearson correlations	Between-day stability of individual physical activity measures was low to moderate (r=-0.23 to 0.53), indicating that when using accelerometry or heart rate telemetry more than 3d of monitoring is needed to assess usual activity.	High	N/NR
40	Kim & Lochbaum (2018, United States)	Children (N=51; 10.3±0.91 yrs; 65% females; Non-Hispanic Black (n=31), Hispanic (n=12), Others (n=8))	3 days	*Polar Active Watch* (non-dominant wrist; Polar Websync Software; 30 sec; Threshold #1: sedentary <2 MET, LPA 2-3.49, MVPA ≥3.5; #2: sedentary: <1.5,	Intensity (Time spent in different intensities (sedentary, LPA, MVPA))	Wearable (ActiGraph GT3X+ (waist), ActiGraph	Bland-Altman analysis; Mean absolute percentage error	The PAW showed moderate convergent validity for sedentary and MVPA minutes against the GT3X+/GT9X accelerometers.	High	N/N

				LPA, 1.5-2.99, MVPA ≥3; #3: sedentary <2, LPA 2.01-3.99, MVPA ≥ 4)		GT9X (wrist))				
42	Klesges & Klesges (1987, United States)	Toddlers (N=28; 2-4 yrs; 46% females; NR)	< 1 day (9 hours)	*Caltrac Personal Act. Computer* (left hip; NR; NR; NR)	Intensity (Physical activity levels)	Observation (Direct)	Spearman correlation; Step-wise multiple regression	Range of correlations .62 to .95; all-day accelerometry correlated with observational instrument (rho=.54, p<.01).	High	N/NR
43	Krishnaveni et al. (2009, India)	Indian children (N=103; 6.6±0.4 yrs; 65% females; NR)	7 days	*ActiGraph AM7164* (right hip; NR; 1 minute; NR)	Intensity (Energy expenditure)	Diary (Reported by parents and teachers)	Bland-Altman analysis; Kappa statistics	Though accelerometer counts correlate with time spent in activity of varying intensity and energy expenditure derived from parent-maintained diaries, wide limits of agreement show that the limitations of accelerometers need to be recognized in interpreting the data that they generate.	High	N/N
48	Mooses et al. (2013, Estonia)	Third grade students (N=174; 9-10 yrs; 50% females; NR)	5 days	*Fitbit Zip* (hip; NR; 1 minute; NR)	Intensity (Time spent in different intensities (sedentary, LPA, MVPA); Steps)	Wearable (ActiGraph GT3x-BT (hip))	Wilcoxon Signed Rank test; Spearman correlations; Bland-Altman analysis	In general, the Fitbit Zip can be considered a relatively accurate device for measuring the number of steps, MVPA and sedentary time in students in a school-setting. However, in segments where sedentary time dominates (e.g. academic classes), a research-grade accelerometer should be preferred.	High	N/N
52	Oliver et al. (2006, New Zealand)	Preeschool children (N=13; 4.1±0.6 yrs; 46% females; NR)	< 1 day (35 min)	*Yamax Digi-Walker SW-200* (hip; NR; NR; NR)	Intensity (Steps)	Observation (Direct)	Regression analyses; Spearman correlations	Limits of agreement and prediction intervals for directly observed step counts were also wide for pedometers, calling into question their acceptability for use with preschoolers.	High	N/NR
54	Pulakka et al. (2013, Malawi)	Toddlers (N=56; 16-18 months; 57% females; NR)	< 1 day (2 hrs)	*ActiGraph GT3X* (right hip; NR; 15 sec; NR)	Intensity (Counts)	Observation (Video)	Receiver operating characteristic (ROC) curve analysis	The accelerometer proved a feasible and valid method of assessing physical activity among Malawian toddlers.	High	N/N
56	Ramirez-Marrero et al. (2004, United States)	African-American children (N=12; 8.1±0.9 yrs; 58% females; African-Amercian (N=12))	7 days	*Tritrac-R3D* (waist; NR; NR; NR); *Yamax SW-200 Digi-Walker* (hip; NR; NR; NR)	Intensity (Physical activity energy expenditure; Steps; Counts)	Doubly labeled water	One-way ANOVA; Simple linear regression; Pearson correlations; Pairwise comparison of the means and the differences	With some limitations the Tritrac and Digiwalker can provide useful and accurate information about PA and EE in 7- to 10-year old children.	High	N/NR
59	Rowlands & Eston (2005, United Kingdom)	Children (N=34; 9.5±0.7 yrs; 50% females; NR)	7 days	*Yamax Digi-Walker DW-200* (waist; NR; NR; NR)	Intensity (Steps)	Wearable (Tritrac-R3D (waist))	Correlations; Sensitivity; Specificity	These pedometer thresholds provide a reasonable estimation when assessment of physical activity intensity is not possible.	High	NR/NR

60	Rowlands et al. (2014, Australia)	Healthy children (N=58; 10.7+0.8 yrs; 47% females; NR)	7 days	*GENEActiv* (wrist, hip; GENEActiv PC software v.2.2; 1 sec; GEN_{HIP}: SED, <3g*s; LIGHT, ≥3g*s to ≤16g*s; MOD, >16g*s to ≤51g*s; VPA, >51g*s; GEN_{WRIST}: SED, <7g*s; LIGHT, ≥7g*s to ≤19g*s; MOD, >19g*s to ≤60g*s; VPA, >51g*s)	Intensity (Time spent in different intensities (sedentary, LPA, MVPA))	Wearable (ActiGraph GT3X+ (hip))	Correlations; Repeated-measures ANOVA; Pairwise comparisons; Linear regression analyses	The assessment of children's activity level, time spent sedentary, and time in MVPA estimated from GENEActiv seems to be comparable with that of the uniaxial ActiGraph.	High	N/N
63	Schneider & Chau (2016, United States)	Students: Cohort 1 (N=25; 12.76±0.72; yrs; 52% females; 32% Hispanic, 24% Non-Hispanic White, 32 % African-American, 12 % Other); Cohort 2 (N=35; 11.15±0.43 yrs; 53% females; 44% Hispanic, 27% Non-Hispanic White, 15% African-American, 12% Other); Cohort 3 (N=27; 12.74±0.52 yrs; 60% females; 41% Hispanic, 30% Non-Hispanic White, 0% African-American, 22 % Other)	7 days	*Fitbit Zip* (waist; Fitabase Software; NR; NR)	Intensity (Steps; Time spent in MVPA)	Wearable (ActiGraph GT3X (waist))	Pearson correlation; Bland-Altman analysis	Fitbit Zip is a reasonable alternative to the ActiGraph for estimating activity among free-living adolescents. However, data from the Fitbit should not be used interchangeably with data from the ActiGraph, as there is a consistent tendency for the Fitbit to overestimate steps in comparison to the ActiGraph.	High	N/N
66	Simunek et al. (2019, Czech Republic)	Middle and High school students (N=185; 15.9±0.9 yrs; 52% females; NR)	7 days	*Garmin Vivofit 1* (non-dominant upper limb; Garmin connect application; NR; NR); *Polar Loop* (wrist; NR; NR; NR)	Intensity (Steps)	Wearable (Yamax Digiwalker SW-701 (hip); ActiGraph GT3X (hip))	Pearson correlation; Mean absolute percentage error; Bland-Altman analysis	Vívofit showed Higher validity than Loop in measuring daily step counts in free-living conditions. Loop appears to overestimate the daily number of steps in individuals who take more steps during a day.	High	N/N
67	Sirard et al. (2017, United States)	Children (N=16; 8.6±1.6 yrs; 50% females; NR)	4 days	*MovBand Model 2* (dominant wrist; device website; NR; proprietary algorithms); *Sqord* (dominant wrist; device website; NR; proprietary algorithms); *Zamzee* (right hip; device website; 10 sec; proprietary algorithms)	Intensity (Steps; Time spent in activities (metrics))	Wearable (ActiGraph GT3X+ (hip))	Pearson correlation; Repeated measures linear models	Across study phases, the SQ demonstrated stronger validity than the MB and ZZ. The validity of youth-oriented activity trackers may directly impact their effectiveness as behavior modification tools, demonstrating a need for more research on such devices.	High	N/N

70	Tanaka et al. (2019, Japan)	Primary school children (Study 1: N=30; Study 2: N=108, 9.3±1.7 yrs; 62% females; NR)	6 days	Study 1: *Kenz Lifecorder* (hip; NR; NR; NR); Study2: *Yamasa EX-200* (hip/pocket pants; NR; NR; NR); *Omron Active style Pro HJA-350IT* (hip; NR; NR; NR)	Intensity (Steps)	Wearable (Study 1: Yamax SW-200 (hip); Study 2: Kenz Lifecorder (hip)	Percentage difference; Mean absolute percent error; Pearson correlation; Bland-Altman analysis	The choice of pedometer had a substantial impact on step counts.	High	N/Y
71	Treuth et al. (2003, United States)	Girls (N=68, 9±0.6 yrs; 100% females, NR)	4 days	*Yamax Digi-Walker SW-200* (left hip; NR; NR)	Intensity (Steps)	Wearable (MTI/CSA accelerometer (hip))	Pearson Correlations; Paired t-test	Validity correlations were significant when more than one day was used.	High	N/NR
73	Van Hoye et al. (2014, France)	Children (N=18; 11.9±1.97 yrs; 59% females; NR)	7 days	*SenseWear Pro 3 Armband* (upper arm; SensewearPro 6.1 software; 1 minute; light 1.50-2.99, moderate 3.00-5.99, vigorous > 6)	Intensity (Steps; Time spent in different intensities (LPA, MVPA))	Wearable (ActiGraph wGT3X (waist))	Pearson correlation; Paired t-tests	Large significant correlations between both accelerometers for number of steps and physically active days, and for time spent in light and MVPA intensity (rs > 0.59, Ps < 0.01).	High	N/N
74	Voss et al. (2017, Canada)	Children with congenital heart disease (N=30; 13±2.2 yrs; 53% females; NR)	7 days	*Fitbit Charge HR* (wrist; online dashboard/Fitabase; 1 minute; proprietary algorithm (≥ 3 MET´s)	Intensity (Steps; Time spent in different intensities (sedentary; MVPA))	Wearable (Actigraph GT3X+/GT9 X (hip/wrist))	Intraclass correlation coefficient; Bland-Altman analysis; mean absolute percent error; Sensitivity, Specificity	Commercial activity trackers provide opportunities to remotely monitor physical activity in children with CHD, but absolute values might differ from accelerometers.	High	N/N
75	Welk et al. (1998, United States)	Children (N=32; 10-12 yrs; NR, NR)	< 1 day (1 hr 10 min)	*Tritrac accelerometer* (NR; NR; NR; NR)	Intensity (Counts)	Observation (Direct)	Pearson correlation; Regression analysis	Collectively, the combination ofTritrac and HR provided little advantage over the assessment provided by either measure alone.	High	N/NR
76	Xi et al. (2019, China)	Children (N=99; 13.0±2.5 yrs; 52% females; Majority (n=91), Minority (n=8))	7 days	*Wristband* (non-dominant wrist; NR; NR; NR)	Intensity (Steps; Physical activity energy expenditure)	Wearable (ActiGraph wGT3X-BT (hip))	Spearmans correlation; Median of absolute percentage error; Bland-Altman analysis	The wristband activity monitor seems to be reliable and valid for measurement of overall children's physical activity, providing a feasible objective method of physical activity surveillance in children.	High	N/N

[1]N= No; [2]NR=Not reported; [3]Y=Yes