BMC Public Health

CrossMark

Reproducibility of domain-specific physical activity over two seasons in children

Eivind Aadland[1*] iD, Lars Bo Andersen[1], Ulf Ekelund[2], Sigmund Alfred Anderssen[1,2] and Geir Kåre Resaland[1,3]

Abstract

Background: Knowledge of the reproducibility of domain-specific accelerometer-determined physical activity (PA) estimates are a prerequisite to conduct high-quality epidemiological studies. The aim of this study was to determine the reproducibility of objectively measured PA level in children during school hours, afternoon hours, weekdays, weekend days, and total leisure time over two different seasons.

Methods: Six hundred seventy six children from the Active Smarter Kids study conducted in Sogn og Fjordane, Norway, were monitored for 7 days by accelerometry (ActiGraph GT3X+) during January–February and April–May 2015. Reproducibility was estimated week-by-week using intra-class correlation (ICC) and Bland-Altman plots with 95% limits of agreement (LoA).

Results: When controlling for season, reliability (ICC) was 0.51–0.66 for a 7-day week, 0.55–0.64 for weekdays, 0.11–0.43 for weekend days, 0.57–0.63 for school hours, 0.42–0.53 for afternoon hours, and 0.42–0.61 for total leisure time. LoA across models approximated a factor of 1.3–2.5 standard deviations of the sample PA levels. 3–6 weeks of monitoring were required to achieve a reliability of 0.80 across all domains but weekend days, which required 5–32 weeks.

Conclusion: Reproducibility of PA during leisure time and weekend days were lower than for school hours and weekdays, and estimates were lower when analyzed using a week-by-week approach over different seasons compared to previous studies relying on a single short monitoring period. To avoid type 2-errors, researchers should consider increasing the monitoring period beyond a single 7-day period in future studies.

Trial Registration: ClinicalTrials.gov, NCT021324947. Registered on 7 April 2014.

Keywords: Test-retest, Reliability, Agreement, Intra-class correlation, Measurement, Accelerometry

Background

There are many unresolved issues regarding data reduction and quality assessment of accelerometry data to determine physical activity (PA) and sedentary behavior (SED). These challenges leads to great variation in procedures used and criteria applied to define a valid measurement [1]. As behavior vary greatly over time, an important aspect of accelerometer measurements is how many days or periods of measurement that need to be included to obtain reliable estimates of habitual activity level. This is particularly true when children live in an area with a significant change in weather during different seasons [2–4]. Measurement error caused by poor reproducibility (amongst other sources of

error) may preclude researchers from arriving at valid conclusions and possibly misinform the society regarding PA as a target for public health initiatives [5].

Most studies in children apply a criterion of a minimum 3 or 4 wear days to constitute a reproducible accelerometer-measurement [1]. Although findings vary between studies in both adults [6–10] and children [11–22], most evidence suggest that a reasonable reliability (i.e., intra-class correlation (ICC)) of ~ 0.70–0.80 are achieved with 3–7 days of monitoring. However, the reproducibility might vary across PA domains, as suggested by a study in adults [7]. Given the importance of evaluating effects of interventions in preschool [23, 24] and school settings [25, 26], especially considering that there might be a reactivity effect for PA [27], it is critical to determine activity level during school hours and leisure time separately. Moreover, many studies target in-school versus out-of-school, or weekday

* Correspondence: eivind.aadland@hvl.no
[1]Department of Sport, Food and Natural Sciences, Faculty of Education, Arts and Sports, Western Norway University of Applied Sciences, Campus Sogndal, Box 133, 6851 Sogndal, Norway
Full list of author information is available at the end of the article

Aadland *et al. BMC Public Health* (2018) 18:821

Page 2 of 9

versus weekend activity patterns [28–31]. The validity of such studies depends on whether the intended associations, patterns or effects can be reliably captured during the applied monitoring period.

Most previous studies have estimated the reliability and the number of days needed based on the Spearman Brown prophecy formula for measurements conducted over a single 7-day period [8, 9, 11, 13–22]. Such study designs have received critique for being likely to underestimate the number of days needed, and they should therefore be interpreted with caution [32–34]. Importantly, such results are in principle only generalizable to the included days, as inclusion of additional days, weeks or seasons will add variability. Few studies have determined the reliability for longer periods, of which all have shown considerable intra-individual variation [34–38]. Reliability has been shown to be ~ 0.70–0.80 with limits of agreement (LoA) ~ 1 standard deviation (SD) for one out of two and three consecutive weeks of measurement in preschool children and adults, respectively [37, 38]. However, poorer estimates are found in studies considering seasonality [34–36], leaving reliability estimates of ~ 0.50 for one week monitoring in children [34, 36]. These findings agrees with studies showing substantial seasonal variation in activity level in children and adolescents [2–4], which are obviously not captured when relying on a single measurement period. Finally, agreement (i.e., LoA and/or standard error of the measurement), which provide researchers a direct quantification of how much outcomes should be expected to vary over time [39–41], has not been reported for domain-specific PA or SED in children.

The aim of the present study was to determine the domain-specific reproducibility of accelerometer-determined PA and SED for one week of measurement obtained during two different seasons separated by 3–4 months in a large sample of children. We hypothesized great variability across weeks and reliability estimates lower than ICC = 0.80 for all accelerometer outcomes, but somewhat better reliability for school hours compared to leisure time.

Methods

Participants

The present analyses are based on data obtained in fifth grade children from the Active Smarter Kids (ASK) cluster-randomized trial, conducted in Norway during 2014–2015 [26, 42]. The main aim of the ASK study was to investigate the effect of school-based PA on academic performance and various health outcomes. Physical activity was measured with accelerometry at baseline (mainly May to June 2014) and follow-up (April to May 2015) in all children, as well as in approximately two-thirds of the children that we invited to complete a mid-term (January to February 2015) measurement. In the present study, we include the mid-term and the

follow-up measurement, to allow for comparison of PA over two different seasons separated by 3–4 months. Additionally, as the intervention was ongoing at both these time-points, and we found no effect of the intervention on PA [26], we included both the intervention and the control groups. We have previously published a detailed description of the study [42], and do only provide a brief overview of the accelerometer handling herein.

Our procedures and methods conform to ethical guidelines defined by the World Medical Association's Declaration of Helsinki and its subsequent revisions. The Regional Committee for Medical Research Ethics approved the study protocol. We obtained written informed consent from each child's parents or legal guardian and from the responsible school authorities prior to all testing. The study is registered in Clinicaltrials.gov with identification number: NCT02132494.

Procedures

Physical activity was measured using the ActiGraph GT3X+ accelerometer (Pensacola, FL, USA) [43]. During both measurements, participants were instructed to wear the accelerometer at all times over 7 consecutive days, except during water activities (swimming, showering) or while sleeping. Units were initialized at a sampling rate of 30 Hz. Files were analyzed at 10 s epochs using the KineSoft analytical software version 3.3.80 (KineSoft, Loughborough, UK). In all analyses, consecutive periods of ≥20 min of zero counts were defined as non-wear time [1, 44]. Results are reported for overall PA level (cpm), as well as minutes per day spent SED (< 100 cpm), in light PA (LPA) (100–2295 cpm), in moderate PA (MPA) (2296–4011 cpm), in vigorous PA (VPA) (≥ 4012 cpm), and in moderate-to-vigorous PA (MVPA) (≥ 2296 cpm), determined using previously established and validated cut points [45, 46].

Results were reported for a 7-day week (mean of up to 7 days providing ≥8 h of wear time), weekdays (mean of up to 5 days providing ≥8 h of wear time), weekend days (mean of up to 2 days providing ≥8 h of wear time), school hours (mean of up to 5 days providing ≥4 h of wear time), afternoon (mean of up to 5 days providing ≥4 h of wear time), and total leisure time (mean of afternoon and weekend days). Data for a full day was restricted to hours 06:00 to 23:59, whereas we defined school hours from 09:00 to 14:00, and afternoon from 15:00 to 23.59. We only included children that provided data for all domains for analysis, that is, ≥ 3 weekdays, school days and afternoons, as well as ≥1 weekend day. We also conducted sensitivity analyses restricted to children providing ≥4 weekdays, school days and afternoons, and 2 weekend days.

Aadland *et al. BMC Public Health* (2018) 18:821

Page 3 of 9

Statistical analyses

Children's characteristics were reported as frequencies, means and SDs. Differences between included and excluded children and differences in PA and SED between measurements was tested using a mixed effect model including random intercepts for children. Wear time was included as a covariate for analyses of PA and SED. Differences in PA and SED between measurements were reported as standardized effect sizes (ES = mean difference between seasons/square root(($SD_{winter}^2 + SD_{spring}^2/2$))).

We estimated reproducibility using a week-by-week approach [37, 38]. Reliability for a single week of measurement (ICC_s) was assessed using variance partitioning applying a one-way random effect model not controlling for season (i.e., determining reliability based on an absolute agreement definition) and a two-way mixed effect model controlling for season (i.e., determining reliability based on a consistency definition) [47]. In the latter model, the variance attributed to season is removed, thus, it determines to what degree children retain their rank between the two time points, without respect to the mean difference between the two measurements. All models were adjusted for wear time. Number of weeks needed to obtain a reliability of 0.80 (N) was estimated using the Spearman Brown prophecy formula (ICC for average measurements [ICC_k]) [6, 47]: N = ICC_t/(1-ICC_t)*[(1-ICC_s)/ICC_s], where N = the number of weeks needed, ICC_t = the desired level of reliability, and ICC_s = the reliability for a single week.

We additionally applied Bland Altman plots to assess agreement, showing the difference between two subsequent weeks as a function of the mean of the two weeks [39]. We calculated 95% LoA from the residual variance (i.e., within-subjects) error term based on the variance partitioning models (LoA = √residual variance *√2*1.96) [41].

All analyses were performed using IBM SPSS v. 23 (IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp., USA). A *p*-value < .05 indicated statistically significant findings.

Results

Participants' characteristics

Of the 1129 included children in the ASK-study, 676 provided accelerometer data at the mid-term and post measurement, of whom 465 children aged 10.4 to 11.6 years (69% of the total sample; 48% boys) provided valid data for analyses of PA (Table 1). There were no differences between the included (*n* = 465) and excluded (664) children in anthropometry (*p* ≥ .064). During school hours, the children included in the present analyses were more SED (mean (95% CI) 3.7 (1.5–5.9) min/day, *p* = .001), and demonstrated lower overall (− 41 (− 64–-18) cpm, p = .001) and intensity-specific PA levels (MPA: − 1.7 (− 2.4–-0.9) min/day; VPA: − 1.3 (− 2.1–-0.6) min/day; MVPA: − 3.0

Table 1 The children's characteristics (*n* = 447–465). Values are mean (SD) if not otherwise stated

	Boys (*n* = 225)	Girls (*n* = 240)	Total (*n* = 465)
Age (years)	10.9 (0.3)	10.9 (0.3)	10.9 (0.3)
Body mass (kg)	39.1 (8.4)	39.3 (8.7)	39.2 (8.5)
Height (cm)	146 (7)	146 (7)	146 (7)
BMI (kg/m^2)	18.2 (2.9)	18.2 (3.1)	18.2 (3.0)
Overweight/obese (%)	17/4	15/4	16/4

BMI Body mass index

(− 4.3–-1.7) min/day; all *p* ≤ .001), compared to the excluded children. Differences were smaller among other domains.

In general, overall PA level (cpm) and intensity-specific PA was significantly higher, and SED was significantly lower, in the spring than in the winter (ES = 0.08–0.75 for the total 7-d week) (Table 2; Additional file 4: Table S1). Differences were modest during school hours (ES = 0.01–0.21), but much greater during the afternoon (ES = 0.27–0.87) and during total leisure time (ES = 0.23–0.91). Differences for weekdays (ES = 0.06–0.67) and weekend days (ES = 0.10–0.60) were similar.

Reproducibility

As PA and SED differed substantially between the two measurements for most outcomes, reliability estimates improved when controlling for season (i.e., applying a consistency definition of reliability) versus when not controlling for season (i.e., applying an absolute agreement definition of reliability). The difference between estimates was greatest for variables having the greatest bias. A 7-d week provided ICC values 0.31–0.66 when not controlling for season, and 0.51–0.66 when controlling for season (Table 3). Reliability estimates for weekdays was very similar to an overall 7-d week, whereas weekend days clearly provided poor reliability estimates whether controlling for season or not (ICC = 0.01–0.43). On the contrary, reliability estimates for school hours was modest, being more or less identical whether controlling for season or not (ICC = 0.56–0.63) (Table 4), due to the small bias. Estimates for afternoon was lower than for school hours, with clear improvements when controlling for season, due to the much larger bias between seasons than for school hours (ICC = 0.15–0.50 when not controlling for season vs. 0.42–0.53 when controlling for season). The total leisure time estimates was similar to the estimates for the afternoon (ICC = 0.14–0.59 and 0.42–0.61 when not controlling and controlling for season, respectively).

When controlling for season, the number of weeks needed to obtain a reliability of 0.80 as estimated by the Spearman Brown prophecy formula was 2.2–3.9 for an overall 7-d week, 2.2–3.2 for weekdays, 5.3–31.4 for

Aadland *et al. BMC Public Health* (2018) 18:821

Page 4 of 9

Table 2 Physical activity level during the winter and spring measurements. Values are mean (SD)

Physical activity level*	7-d Week	Weekdays	Weekend	School	Afternoon	Leisure
	Winter					
Wear days (days/week)	6.5 (0.8)	4.7 (0.6)	1.8 (0.4)	4.5 (0.7)	4.6 (0.6)	6.4 (0.8)
Wear minutes (min/day)	778 (46)	807 (49)	700 (83)	295 (6)	382 (39)	470 (44)
Overall PA (cpm)	481 (125)	510 (134)	406 (163)	582 (167)	438 (167)	429 (147)
SED (min/day)	508 (50)	521 (54)	472 (77)	183 (18)	259 (40)	318 (43)
LPA (min/day)	215 (33)	225 (34)	190 (47)	85 (13)	99 (21)	124 (25)
MPA (min/day)	35 (10)	38 (12)	25 (13)	16 (5)	14 (6)	17 (7)
VPA (min/day)	18 (9)	21 (11)	11 (9)	13 (6)	8 (6)	9 (6)
MVPA (min/day)	53 (18)	59 (20)	36 (21)	26 (10)	22 (11)	26 (12)
	Spring					
Wear days (days/week)	6.4 (0.6)	4.6 (0.6)	1.8 (0.4)	4.4 (0.7)	4.6 (0.6)	6.4 (0.8)
Wear minutes (min/day)	789 (48)	812 (48)	732 (90)	295 (6)	388 (36)	485 (43)
Overall PA (cpm)	617 (204)	630 (200)	580 (368)	621 (201)	667 (295)	643 (268)
SED (min/day)	499 (54)	509 (56)	475 (87)	183 (18)	247 (39)	311 (46)
LPA (min/day)	221 (33)	228 (33)	203 (50)	83 (13)	107 (21)	134 (25)
MPA (min/day)	39 (12)	42 (13)	31 (16)	17 (6)	18 (7)	22 (8)
VPA (min/day)	26 (14)	29 (15)	20 (19)	13 (7)	15 (9)	16 (10)
MVPA (min/day)	66 (24)	71 (25)	51 (32)	28 (11)	33 (15)	38 (17)

PA Physical activity; *cpm* Counts per minute; *SED* Sedentary time; *LPA* Light physical activity; *MPA* Moderate physical activity; *VPA* Vigorous physical activity; *MVPA* Moderate-to-vigorous physical activity; all results are based on n = 465 children

Table 3 The reproducibility of different outcome variables for one out of two weeks of measurement for a 7-d week, weekdays and weekend days, applying a ≥ 3 weekdays and ≥ 1 weekend day wear time criterion (*n* = 465 (69%) children)

	7-day Week			Weekdays			Weekend days		
	ICC$_s$	LoA	N	ICC$_s$	LoA	N	ICC$_s$	LoA	N
	Not corrected for season (absolute agreement definition)								
Overall PA (cpm)	0.31	421	9.1	0.38	395	6.6	0.01	818	291
SED (min/day)	0.60	74.2	2.6	0.59	78.5	2.8	0.32	134	8.4
LPA (min/day)	0.66	49.5	2.1	0.62	53.3	2.5	0.43	89.2	5.4
MPA (min/day)	0.55	21.2	3.3	0.54	23.1	3.4	0.21	36.0	14.8
VPA (min/day)	0.43	25.9	5.2	0.47	27.4	4.5	0.09	41.6	38.3
MVPA (min/day)	0.52	42.1	3.7	0.54	44.9	3.5	0.16	70.1	21.1
	Corrected for season (consistency definition)								
Overall PA (cpm)	0.51	329	3.9	0.55	316	3.2	0.11	742	31.4
SED (min/day)	0.65	68.3	2.1	0.63	73.5	2.4	0.35	129	7.3
LPA (min/day)	0.66	49.2	2.1	0.62	53.2	2.5	0.43	88.8	5.3
MPA (min/day)	0.59	19.8	2.8	0.58	21.8	2.9	0.24	35.0	12.5
VPA (min/day)	0.61	20.4	2.6	0.62	22.2	2.5	0.19	37.9	17.5
MVPA (min/day)	0.64	34.8	2.2	0.64	38.1	2.2	0.23	65.2	13.5

PA Physical activity; *cpm* Counts per minute; *SED* Sedentary time; *LPA* Light physical activity; *MPA* Moderate physical activity; *VPA* Vigorous physical activity; *MVPA* Moderate-to-vigorous physical activity; *ICC$_s$* Intra-class correlation for a single week of measurement adjusted for wear time; *N* Number of weeks needed to achieve a ICC = 0.80; *LoA* 95% Limits of agreement. Estimates not corrected for season are derived from one-way random effect models (i.e., using an absolute agreement definition of reproducibility), whereas models corrected for season are derived from two-way mixed effect models (i.e., using a consistency definition of reproducibility)

Aadland *et al. BMC Public Health* (2018) 18:821

Page 5 of 9

Table 4 The reproducibility of different outcome variables for one out of two weeks of measurement for school hours, afternoon and total leisure time, applying a ≥ 3 weekdays and ≥ 1 weekend day wear time criterion (*n* = 465 (69%) children)

	School hours			Afternoon hours			Leisure time*		
	ICC_s	LoA	N	ICC_s	LoA	N	ICC_s	LoA	N
	Not corrected for season (absolute agreement definition)								
Overall PA (cpm)	0.56	343	3.2	0.15	679	22.8	0.14	621	24.3
SED (min/day)	0.63	30.0	2.4	0.38	63.3	6.5	0.47	66.1	4.6
LPA (min/day)	0.57	22.9	3.0	0.50	38.8	4.0	0.59	41.1	2.8
MPA (min/day)	0.57	10.0	3.0	0.30	16.6	9.3	0.35	17.7	7.3
VPA (min/day)	0.57	12.1	3.0	0.26	19.9	11.6	0.24	21.7	12.5
MVPA (min/day)	0.60	18.3	2.7	0.29	33.5	9.9	0.31	36.0	9.0
	Corrected for season (consistency definition)								
Overall PA (cpm)	0.57	335	3.0	0.42	507	5.6	0.42	462	5.5
SED (min/day)	0.63	30.1	2.4	0.51	54.1	3.9	0.57	57.1	3.0
LPA (min/day)	0.58	22.6	2.9	0.53	37.0	3.5	0.61	39.6	2.5
MPA (min/day)	0.57	10.0	3.0	0.43	14.3	5.2	0.46	15.5	4.7
VPA (min/day)	0.58	12.0	2.9	0.47	15.5	4.5	0.46	16.8	4.6
MVPA (min/day)	0.61	17.8	2.5	0.50	26.1	4.0	0.51	28.3	3.9

PA Physical activity; *cpm* Counts per minute; *SED* Sedentary time; *LPA* Light physical activity; *MPA* Moderate physical activity; *VPA* Vigorous physical activity; *MVPA* Moderate-to-vigorous physical activity; *ICC_s* Intra-class correlation for a single week of measurement adjusted for wear time; *N* Number of weeks needed to achieve a ICC = 0.80; *LoA* = 95% limits of agreement; *Afternoon + Weekend. Estimates not corrected for season are derived from one-way random effect models (i.e., using an absolute agreement definition of reproducibility), whereas models corrected for season are derived from two-way mixed effect models (i.e., using a consistency definition of reproducibility)

weekend days, 2.4–3.0 for school hours, 3.5–5.6 for afternoon, and 2.5–5.5 for total leisure time. A substantial intra-individual variation over time for all outcomes is also indicated by the LoA (Table 3; Table 4; Additional file 1: Figure S1, Additional file 2: Figure S2 and Additional file 3: Figure S3), which approximated factors of 1.3–1.8, 1.3–1.8, 1.6–2.5, 1.7–1.8, 1.4–2.0, 1.3–1.9 SDs of the sample PA levels, respectively. Overall PA level (cpm) was the least reproducible outcome across models. In general, we found minor improvements in reproducibility when the analyses were restricted to those children (*n* = 257, 38%) providing data for ≥4 weekdays and 2 weekend days (Additional file 5: Table S2).

Discussion

The aim of the present study was to determine domain-specific reproducibility of accelerometer-determined PA and SED in children over two separate weeks of monitoring undertaken 3–4 months apart. Our results suggest that one week of accelerometer monitoring have poorer reliability than suggested by most previous studies that have relied on a single monitoring period. Further, our findings revealed that reliability was superior during school hours and weekdays as compared to leisure time.

Previous studies conclude that 3–7 monitoring days provide reliable estimates of PA and SED in children [11–22]. These studies have estimated reliability based on day-by-day analyses, typically using a single 7-day monitoring period. In contrast, but consistent with

studies including measurements over several seasons [34–36], our findings shows that longer and/or a higher number of monitoring periods are required to estimate PA and SED with acceptable confidence. Mattocks [36] determined overall PA, MVPA and SED over four 7-day periods over approximately one year using the Actigraph 7164 accelerometer in 11–12-year-old children. The ICC for one single period of measurement varied from 0.45 to 0.59 across outcome variables. Wickel & Welk [34] found an ICC of 0.46 for one out of three 7-day periods to assess steps for the Digiwalker pedometer in 80 children aged ~ 10 years. The present findings along with previous findings may question the validity of one week of measurement to determine children's habitual activity level.

While levels of PA and SED during weekdays and school hours can be estimated with a reliability of 0.80 across 2–3 week-long monitoring periods, 3–6 weeks of measurement are required to achieve this level of precision for afternoon hours and for total leisure time. This finding is consistent with our hypothesis and a previous study in adults [7]. As variation in behavior are restricted by the school curriculum during school hours, the higher reliability compared to leisure time is expected. During school hours, students follow the same schedule, including a roughly similar time spent in for example physical education and recess, during both monitoring periods. Thus, although some variation in levels and types of activities could be expected also

Aadland *et al. BMC Public Health* (2018) 18:821

Page 6 of 9

during the school day, this variation would be restricted by the applied teaching methods and established curriculum. On the contrary, our findings indicate that children vary their activity levels greatly across season during leisure time, likely according to climate, weather, and daylight. The present study were collected in the Sogn og Fjordane County in western Norway, and we believe several factors might have caused the great variation in leisure time activity level. In the winter (January–February), daylight fade after hours 15:00–17:00, whereas the spring-measurement (April–May) was conducted when there are daylight until bedtime, giving the children a greatly different opportunity to be active outside. Moreover, the weather can be very different, inviting children to play outside in the spring, whereas they are prone to spend their time inside in the winter. There are also large geographic differences between areas. The coast have a mild and wet climate throughout the year, whereas the inland and mountain areas can have a cold winter and generally less precipitation. Thus, children are likely to spend their time in different modes of activity in different locations across the monitoring periods. Because many movement patterns (e.g., swimming, cycling, and skiing) might be poorly captured by accelerometers [48], such variation in preferred activities over time have the potential to greatly influence stability of accelerometer-derived PA and SED levels during leisure time. According to our findings, these (amongst other) sources of variability influence the weekend days the most, for which we found that 5–32 weeks of measurement was required to achieve an ICC = 0.80. Thus, our results suggest that PA and SED on weekend days, whether controlled for season or not, cannot be measured reliably using a feasible protocol. Yet, it should be kept in mind that the comparison of weekend days only include 1 or 2 days, compared to afternoons including 3, 4, or 5 days, thus, a different precision would be expected, despite both domains being leisure time. Finally, the difference in variance between the two monitoring periods (Table 2) could be an explanation for poor to modest reliability estimates, as the statistical model assumes compound symmetry and the ICC are sensitive to asymmetry [47].

As noise in exposure (x) variables will lead to attenuation of regression coefficients (regression dilution bias), and noise in outcome (y) variables will increase standard errors [5], unreliable measures weaken researchers ability to make valid conclusions. In epidemiology, researchers are in general interested in the long-term habitual PA level, not the very recent days. For example, when evaluating school-interventions, we assume a five-day monitoring period provide a reasonable estimate of true PA or SED. In the school context, our results provide support for modest errors, despite estimates being poorer than suggested in many previous studies. On

the contrary, studies evaluating home-based interventions or studies investigating patterns of weekday versus weekend, or in-school and out-of-school activity levels are prone to type 2-errors, if relying on a 7-day monitoring period that provide an insufficient snapshot of children's habitual activity level. Consequently, such studies must be sufficiently powered. Sample size calculations for designs with repeated measurements normally correct expected SDs for the expected correlations between the baseline and follow-up measurements (i.e., studies use SD for change rather than the cross-sectional SD), thus, the sample size needed is less than for performing cross-sectional analyses if the pre-post correlation is > 0.50. Yet, this benefit is not achievable for PA outcomes during leisure time given the present results. These findings are therefore important for informing study designs and sample size calculations, as well as for interpretation of study findings.

Although an increased accelerometer monitoring length improve reproducibility and thus improve validity of study conclusions, given that accelerometry is a valid measure of the outcome, the burden for study participants should be kept minimal to maximize response rate and compliance. Yet, we have previously performed 2- and 3-week monitoring protocols in preschool children and adults, respectively, without any problems regarding compliance [37, 38]. More recently, we have also successfully performed a 2-week monitoring protocol in larger samples of children, adults and older people, demonstrating this protocol's acceptance in various context. Yet, performing measurements over separate as opposed to consecutive periods might pose an increased burden for participants, as well as for researchers. Importantly, the required monitoring volume is a matter of the research question posed, as population-estimates on a group level requires less precision than individual-level estimates used for correlational analyses [33].

Strengths and limitations

The main strength of the present study is the inclusion of a large and representative sample of children. As reliability estimates depends on the sample variation [39–41], the reliability estimates presented herein should be generalizable to other large-scale population-based studies. Another strength is inclusion of measurements conducted 3–4 months apart, during two different seasons, as has only been analyzed in a few previous studies in children [34, 36]. This approach extends findings from previous studies in children that have mainly estimated reproducibility over a single short monitoring period [11, 13–22]. A limitation, though, is the inclusion of only two weeks and two seasons, as inclusion of more observations probably would introduce more variability and lead to more conservative reproducibility estimates [34, 36]. Moreover, Norway has profound seasonal differences in weather conditions. This characteristic might limit

Aadland *et al. BMC Public Health* (2018) 18:821

Page 7 of 9

generalizability to areas with less pronounced seasonality. Finally, the inclusion of the intervention group in the current analyses might have caused additional variation to the data, as the intervention group participants could be expected to change their PA level over time. Yet, the intervention was ongoing during both measurements, there was no effect of the intervention on PA levels [26], and reproducibility estimates differed marginally and non-systematically between the intervention and control groups. The maximum ICC difference between groups across the variables analyzed were 0.09 for a 7-day week, 0.06 for weekdays, 0.14 for weekend days, 0.06 for school hours, 0.07 for afternoon hours, and 0.09 for total leisure time (results not shown).

Conclusion

We conclude that the reproducibility for one week of accelerometer monitoring is poor to modest across different domains of PA when seasonal differences is considered. Reliability for a 7-day period was lower than in most previous studies relying on a single monitoring period, and reliability for leisure time and weekend days was lower than for school hours and week days. Longer or repeated measurement periods are favourable compared with one single 7-day period when assessing PA and SED by accelerometry, as this will reduce the possibility of type 2-errors in future studies.

Additional files

Additional file 1: Figure S1. Bland Altman plots of agreement for overall physical activity level (cpm) for different domains over two consecutive weeks of measurement. Bland Altman plots (mean of two weeks of measurement on the x-axis versus the difference between them on the y-axis) for a 7-day week, weekdays, weekend days, school hours, afternoon and total leisure time. All results are based on $n = 465$ children. The full line is the bias between weeks, whereas the dotted lines are 95% limits of agreement. (TIFF 4673 kb)

Additional file 2: Figure S2. Bland Altman plots of agreement for sedentary time (min/day) for different domains over two consecutive weeks of measurement. Bland Altman plots (mean of two weeks of measurement on the x-axis versus the difference between them on the y-axis) for a 7-day week, weekdays, weekend days, school hours, afternoon and total leisure time. All results are based on $n = 465$ children. The full line is the bias between weeks, whereas the dotted lines are 95% limits of agreement. Please be aware that variability is not directly comparable between full days and part of days, due to different wear time. (TIFF 4520 kb)

Additional file 3: Figure S3. Bland Altman plots of agreement for moderate-to-vigorous physical activity (min/day) for different domains over two consecutive weeks of measurement. Bland Altman plots (mean of two weeks of measurement on the x-axis versus the difference between them on the y-axis) for a 7-day week, weekdays, weekend days, school hours, afternoon and total leisure time. All results are based on n = 465 children. The full line is the bias between weeks, whereas the dotted lines are 95% limits of agreement. Please be aware that variability is not directly comparable between full days and part of days, due to different wear time. (TIFF 5467 kb)

Additional file 4: Table S1. Differences (spring – winter) in physical activity level between seasons across all physical activity outcomes. (DOCX 16 kb)

Additional file 5: Table S2a. The reliability for different outcome variables for one out of two weeks of measurement for a 7-d week, weekdays and weekend days, applying a ≥ 4 weekdays and 2 weekend days wear time criterion ($n = 257$ (38%) children). **Table S2b.** The reliability for different outcome variables for one out of two weeks of measurement for school hours, afternoon and total leisure time, applying a ≥ 4 weekdays and 2 weekend days wear time criterion (n = 257 (38%) children). (DOCX 20 kb)

Abbreviations
CI: Confidence interval; CPM: Counts per minute; ICC: Intra-class correlation; ICC_k: Intra-class correlation for average measurements; ICC_s: Intra-class correlation for a single week of measurement; ICC_t: Desired intra-class correlation; LoA: Limits of agreement; LPA: Light physical activity; MPA: Moderate physical activity; MVPA: Moderate-to-vigorous physical activity; n: Number of participants; N: Number of weeks needed to achieve a ICC = 0.80; PA: Physical activity; SD: Standard deviation; SED: Sedentary time; VPA: Vigorous physical activity

Availability of data and materials
The datasets used for the current study are available from the corresponding author on reasonable request.

Authors' contributions
SAA, GKR, LBA and UE obtained funding for the study. EAA, GKR and SAA designed the study. EAA and GKR performed the data collection. EAA analyzed the data and wrote the manuscript draft. All authors discussed the interpretation of the results, and read and approved the final manuscript.

Ethics approval and consent to participate
The South-East Regional Committee for Medical Research Ethics approved the study protocol (reference number 2013/1893). We obtained written informed consent from each child's parents or legal guardian and from the responsible school authorities prior to all testing.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Sport, Food and Natural Sciences, Faculty of Education, Arts and Sports, Western Norway University of Applied Sciences, Campus Sogndal, Box 133, 6851 Sogndal, Norway. [2]Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway. [3]Center for Health Research, Førde Central Hospital, Førde, Norway.

References
1. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health. 2013;10(3):437–50.

Aadland *et al. BMC Public Health* (2018) 18:821

Page 8 of 9

2. Atkin AJ, Sharp SJ, Harrison F, Brage S, Van Sluijs EMF. Seasonal variation in children's physical activity and sedentary time. Med Sci Sports Exerc. 2016; 48(3):449–56. https://doi.org/10.1249/mss.0000000000000786.

3. Gracia-Marco L, Ortega FB, Ruiz JR, Williams CA, Hagstromer M, Manios Y, et al. Seasonal variation in physical activity and sedentary time in different European regions. The HELENA study. J Sports Sci. 2013;31(16):1831–40. https://doi.org/10.1080/02640414.2013.803595.

4. Ridgers ND, Salmon J, Timperio A. Too hot to move? Objectively assessed seasonal changes in Australian children's physical activity. Int J Behav Nutr Phys Act. 2015;12:77. https://doi.org/10.1186/s12966-015-0245-x.

5. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. Br Med J. 2010;340:2289. https://doi.org/10.1136/bmj.c2289.

6. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc. 2005;37(11):S531–S43. https://doi.org/10.1249/01.mss.0000185657.86065.98.

7. Pedersen ESL, Danquah IH, Petersen CB, Tolstrup JS. Intra-individual variability in day-to-day and month-to-month measurements of physical activity and sedentary behaviour at work and in leisure-time among Danish adults. BMC Public Health. 2016;16:1222. https://doi.org/10.1186/s12889-016-3890-3.

8. Jerome GJ, Young DR, Laferriere D, Chen CH, Vollmer WM. Reliability of RT3 accelerometers among overweight and obese adults. Med Sci Sports Exerc. 2009;41(1):110–4. https://doi.org/10.1249/MSS.0b013e3181846cd8.

9. Coleman KJ, Epstein LH. Application of generalizability theory to measurement of activity in males who are not regularly active: a preliminary report. Res Q Exerc Sport. 1998;69(1):58–63.

10. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act. 2011;8:62. https://doi.org/10.1186/1479-5868-8-62.

11. Basterfield L, Adamson AJ, Pearce MS, Reilly JJ. Stability of habitual physical activity and sedentary behavior monitoring by accelerometry in 6-to 8-year-olds. J Phys Act Health. 2011;8(4):543–7.

12. Addy CL, Trilk JL, Dowda M, Byun W, Pate RR. Assessing preschool children's physical activity: how many days of accelerometry measurement. Pediatr Exerc Sci. 2014;26(1):103–9. https://doi.org/10.1123/pes.2013-0021.

13. Hinkley T, O'Connell E, Okely AD, Crawford D, Hesketh K, Salmon J. Assessing volume of accelerometry data for reliability in preschool children. Med Sci Sports Exerc. 2012;44(12):2436–41. https://doi.org/10.1249/MSS.0b013e3182661478.

14. Hislop J, Law J, Rush R, Grainger A, Bulley C, Reilly JJ, et al. An investigation into the minimum accelerometry wear time for reliable estimates of habitual physical activity and definition of a standard measurement day in pre-school children. Physiol Meas. 2014;35(11):2213–28. https://doi.org/10.1088/0967-3334/35/11/2213.

15. Penpraze V, Reilly JJ, MacLean CM, Montgomery C, Kelly LA, Paton JY, et al. Monitoring of physical activity in young children: how much is enough? Pediatr Exerc Sci. 2006;18(4):483–91.

16. Rich C, Geraci M, Griffiths L, Sera F, Dezateux C, Cortina-Borja M. Quality control methods in accelerometer data processing: defining minimum wear time. PLoS One. 2013;8(6):67206. https://doi.org/10.1371/journal.pone.0067206.

17. Murray DM, Catellier DJ, Hannan PJ, Treuth MS, Stevens J, Schmitz KH, et al. School-level intraclass correlation for physical activity in adolescent girls. Med Sci Sports Exerc. 2004;36(5):876–82. https://doi.org/10.1249/01.mss.0000126806.72453.1c.

18. Treuth MS, Sherwood NE, Butte NF, McClanahan B, Obarzanek E, Zhou A, et al. Validity and reliability of activity measures in African-American girls for GEMS. Med Sci Sports Exerc. 2003;35(3):532–9. https://doi.org/10.1249/01.mss.0000053702.03884.3f.

19. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? Med Sci Sports Exerc. 2000;32(2):426–31. https://doi.org/10.1097/00005768-200002000-00025.

20. Chinapaw MJM, de Niet M, Verloigne M, De Bourdeaudhuij I, Brug J, Altenburg TM. From sedentary time to sedentary patterns: accelerometer data reduction decisions in youth. PLoS One. 2014;9(11):111205. https://doi.org/10.1371/journal.pone.0111205.

21. Kang M, Bjornson K, Barreira TV, Ragan BG, Song K. The minimum number of days required to establish reliable physical activity estimates in children aged 2-15 years. Physiol Meas. 2014;35(11):2229–37. https://doi.org/10.1088/0967-3334/35/11/2229.

22. Ojiambo R, Cuthill R, Budd H, Konstabel K, Casajus JA, Gonzalez-Aguero A, et al. Impact of methodological decisions on accelerometer outcome variables in young children. Int J Obes. 2011;35:S98–S103. https://doi.org/10.1038/ijo.2011.40.

23. Pate RR, Brown WH, Pfeiffer KA, Howie EK, Saunders RP, Addy CL, et al. An intervention to increase physical activity in children a randomized controlled trial with 4-year-olds in preschools. Am J Prev Med. 2016;51(1):12–22. https://doi.org/10.1016/j.amepre.2015.12.003.

24. Jones RA, Okely AD, Hinkley T, Batterham M, Burke C. Promoting gross motor skills and physical activity in childcare: a translational randomized controlled trial. J Sci Med Sport. 2016;19(9):744–9. https://doi.org/10.1016/j.jsams.2015.10.006.

25. Kriemler S, Zahner L, Schindler C, Meyer U, Hartmann T, Hebestreit H, et al. Effect of school based physical activity programme (KISS) on fitness and adiposity in primary schoolchildren: cluster randomised controlled trial. BMJ. 2010;340:c785. https://doi.org/10.1136/bmj.c785.

26. Resaland GK, Aadland E, Moe VF, Aadland KN, Skrede T, Stavnsbo M, et al. Effects of physical activity on schoolchildren's academic performance: the active smarter kids (ASK) cluster-randomized controlled trial. Prev Med. 2016; 91:322–8. https://doi.org/10.1016/j.ypmed.2016.09.005.

27. Gomersall SR, Rowlands AV, English C, Maher C, Olds TS. The activitystat hypothesis the concept, the evidence and the methodologies. Sports Med. 2013;43(2):135–49. https://doi.org/10.1007/s40279-012-0008-7.

28. Van Cauwenberghe E, Jones RA, Hinkley T, Crawford D, Okely AD. Patterns of physical activity and sedentary behaviour in preschool children. Int J Behav Nutr Phys Act. 2012;9:138. https://doi.org/10.1186/1479-5868-9-138.

29. O'Neill JR, Pfeiffer KA, Dowda M, Pate RR. In-school and out-of-school physical activity in preschool children. J Phys Act Health. 2016;13(6):606–10. https://doi.org/10.1123/jpah.2015-0245.

30. Nilsson A, Anderssen SA, Andersen LB, Froberg K, Riddoch C, Sardinha LB, et al. Between- and within-day variability in physical activity and inactivity in 9-and 15-year-old European children. Scand J Med Sci Sports. 2009;19(1):10–8. https://doi.org/10.1111/j.1600-0838.2007.00762.x.

31. De Meester F, Van Dyck D, De Bourdeaudhuij I, Deforche B, Cardon G. Changes in physical activity during the transition from primary to secondary school in Belgian children: what is the role of the school environment? BMC Public Health. 2014;14:261. https://doi.org/10.1186/1471-2458-14-261.

32. Baranowski T, Masse LC, Ragan B, Welk G. How many days was that? We're still not sure, but we're asking the question better. Med Sci Sports Exerc. 2008;40(7):S544–S9. https://doi.org/10.1249/MSS.0b013e31817c6651.

33. Matthews CE, Hagstromer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc. 2012;44:S68–76. https://doi.org/10.1249/MSS.0b013e3182399e5b.

34. Wickel EE, Welk GJ. Applying generalizability theory to estimate habitual activity levels. Med Sci Sports Exerc. 2010;42(8):1528–34. https://doi.org/10.1249/MSS.0b013e3181d107c4.

35. Levin S, Jacobs DR, Ainsworth BE, Richardson MT, Leon AS. Intra-individual variation and estimates of usual physical activity. Ann Epidemiol. 1999;9(8):481–8.

36. Mattocks C, Leary S, Ness A, Deere K, Saunders J, Kirkby J, et al. Intraindividual variation of objectively measured physical activity in children. Med Sci Sports Exerc. 2007;39(4):622–9. https://doi.org/10.1249/mss.0b013e318030631b.

37. Aadland E, Johannessen K. Agreement of objectively measured physical activity and sedentary time in preschool children. Preventive Med Rep. 2015;2:635–9.

38. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One. 2015;10(7):1–13. https://doi.org/10.1371/journal.pone.0133296.

39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10.

40. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15.

41. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

42. Resaland GK, Moe VF, Aadland E, Steene-Johannessen J, Glosvik Ø, Andersen JR, et al. Active smarter kids (ASK): rationale and design of a cluster-randomized controlled trial investigating the effects of daily physical activity on children's academic performance and risk factors for non-communicable diseases. BMC Public Health. 2015;15:709. https://doi.org/10.1186/s12889-015-2049-y.

43. John D, Freedson P. ActiGraph and Actical physical activity monitors: a peek under the hood. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S86–S9.

Aadland *et al. BMC Public Health* (2018) 18:821

Page 9 of 9

44. Esliger DW, Copeland JL, Barnes JD, Tremblay MS. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. J Phys Act Health. 2005;2(3):366.
45. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci. 2008;26(14): 1557–65. https://doi.org/10.1080/02640410802334196.
46. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc. 2011;43(7):1360–8. https://doi.org/10.1249/MSS.0b013e318206476e.
47. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46. https://doi.org/10.1037/1082-989x.1.4.390.
48. Tarp J, Andersen LB. Ostergaard L. quantification of underestimation of physical activity during cycling to school when using accelerometry. J Phys Act Health. 2015;12(5):701–7. https://doi.org/10.1123/\2013-0212.