Willems *et al. BMC Public Health* (2024) 24:3607
https://doi.org/10.1186/s12889-024-21139-8

BMC Public Health

RESEARCH

Open Access

Harmonizing measurement tools: examining the concurrent validity of the Daily Activity Behaviors Questionnaire compared to the ActiGraph to assess 24-hour movement behaviors among adults

Iris Willems[1,2], Vera Verbestel[3,4], Dorothea Dumuid[5], Patrick Calders[1], Bruno Lapauw[6] and Marieke De Craemer[1*]

Abstract

Purpose An accurate assessment of time spent in 24-hour movement behaviors (24 h-MBs) is crucial in exploring health related associations. This study aims to evaluate the concurrent validity of the Daily Activity Behavior Questionnaire (DABQ) compared to the ActiGraph using absolute and relative indicators of validity.

Methods This cross-sectional observational study included 105 adults (45 ± 13 y/o, 54% female). Participants wore an ActiGraph during seven consecutive days followed by filling in the DABQ recalling the past seven days. Intraclass correlations (95% confidence intervals), Bland-Altman plots, Spearman's correlations and the magnitude of error were calculated to estimate the absolute agreement and validity. Interaction effects between sociodemographic variables and the measurement methods were explored in mixed models. All analyses were compared by four commonly used data processing methods for ActiGraph data (cut-points and data reduction method-specific).

Results Moderate absolute agreement (ICC = 0.56) and validity (rho_{sleep}=0.58) was found for sleep comparing the DABQ with the ActiGraph. Time spent in sedentary behavior (SB), light physical activity (LPA) and moderate-to-vigorous physical activity (MVPA) showed poor absolute agreement (ICC_{SB}: 0.01–0.38, ICC_{LPA}: 0.00-0.31; ICC_{MVPA}: 0.23–0.30) and validity (rho_{SB}: 0.01–0.43, rho_{LPA}: 0.10–0.46; rho_{MVPA}: 0.38–0.44) comparing the DABQ with the ActiGraph. The Ranges in ICC and Spearmans' rho include the comparison between the four data processing methods. A significant interaction was found between the measurement method and educational level ($p < 0.001$), in specific for sleep, SB and LPA.

Conclusion Compared to the ActiGraph, the DABQ showed accurate time-use estimates for sleep but presented poor to moderate evidence of validity regarding SB, LPA and MVPA. This was shown in underestimations regarding SB and MVPA, and overestimations regarding LPA. However, educational level and data processing methods contributed to these variations.

*Correspondence:
Marieke De Craemer
Marieke.decraemer@ugent.be

Full list of author information is available at the end of the article

Keywords 24-hour movement behaviors, Questionnaire, Accelerometry, Time use epidemiology, Physical activity, Sedentary behavior, Sleep

Background

Time-use epidemiology is a rapidly growing research area that aims to understand how individuals allocate their time to various activities throughout the day [1]. The accurate assessment of daily activity behaviors, such as sleep, sedentary behavior (SB), light physical activity (LPA), and moderate-to-vigorous physical activity (MVPA), is crucial for studying the associations between activity patterns and health outcomes [1]. Device-based measurement (e.g. accelerometers) is the preferred method to objectively quantify individuals' activity levels as devices can be worn over a 24-hour day. However, as these devices can be cost-prohibitive for large studies, researchers often rely on self-report questionnaires [2–4]. Although questionnaires are known for their self-report bias (e.g. recall bias, social desirability), they have advantages of being easy to use, cheap, and able to assess contextual information related to the daily activities [5]. Numerous questionnaires already exist to quantify individuals' behaviors but they are often focused on single movement behaviors without including the bigger picture of movement behaviors across the whole 24h day [3, 4, 6].

A recently developed questionnaire known as the Daily Activity Behavior Questionnaire (DABQ) can be used to measure all 24-hour movement behaviors (24 h-MBs) using a single measurement tool [7]. Validation of the DABQ against the ActivPAL accelerometer showed moderate absolute agreement for sleep duration (ICC 0.6), albeit with lower agreement for SB, LPA, and MVPA estimates (ICC range 0.22–0.47) [7]. The ActivPAL is a thigh-worn accelerometer which is known to provide valid measures of posture and postural transitions (i.e. lying, sitting, standing and stepping) [8–10]. Nevertheless, hip-worn accelerometers are also commonly used to measure time spent in different activity intensities within the context of 24 h-MBs [8, 11]. Previous research has already explored the comparability of the ActiGraph device with single movement questionnaires revealing weak to moderate correlations of self-reported PA intensities (r=0.1–0.6), underestimation of self-reported SB (±2 h/day), and an overestimation of self-reported sleep duration (±1 h/day) [5, 6, 12]. However, it has never been investigated if the 24 h-MBs composition retrieved from the DABQ questionnaire is valid against the ActiGraph accelerometer.

This study will be the first to investigate the absolute agreement and concurrent validity of the DABQ compared with the ActiGraph to measure 24 h-MBs. Moreover, raw accelerometer data of the ActiGraph will be analyzed by four commonly data processing methods for accelerometry (cut-points and data reduction method-specific) to provide a comprehensive overview of current analysis techniques (cut-points by Hildebrand et al. 2014,2017, Vähä Ypyä et al. 2018, 2023, Troiano et al. 2008 and Sasaki et al. 2011) [13–19]. In addition, this study aims to explore how participant characteristics potentially moderate the differences between measurement methods, which is essential to address potential biases in reporting movement behaviors. Factors such as age, sex, and educational level can significantly influence the accuracy of self-reported behaviors [5, 20]. By understanding these sociodemographic influences, we can better interpret discrepancies between questionnaire- and accelerometer-assessed movement behaviors.

Methods

Participants and procedure

This cross-sectional observational study included a convenience sample of adults between 18 and 65 years old. Adults were excluded when working night shifts or when having physical (e.g. amputations, paralysis, recovering from stroke, osteoarthritis), cognitive (e.g. dementia, psychological disorders) and major medical (e.g. Chronic respiratory diseases, heart failure, cardiovascular diseases) conditions that obstruct daily functioning. This study was approved by the ethical committee of Ghent University Hospital and all participants provided a written informed consent prior to the study (ONZ-2023-0384). Participants were visited at home between October 2023 and December 2023. During this home visit, participants were provided with an accelerometer and two questionnaires namely a general information questionnaire and the DABQ, both to be filled in after wearing the accelerometer for seven consecutive days.

Device-measured 24 h-MBs

Data on 24 h-MBs were collected by the tri-axial ActiGraph wGT3x-BT accelerometer. Participants were instructed to wear the device for seven consecutive days [11]. During the day, the device was worn on the right hip [11]. At night, the device was switched to the non-dominant wrist [21]. Participants filled in a diary in which they reported wake up and go to bed times as well as times and reasons for device removal, like water-based activities (e.g. swimming, showering), or activities where wearing the device is not permitted (e.g. contact sports) [11]. The accelerometer was initialized via Actilife software (version) at a frequency of 100 Hz and downloaded at 60 s epochs [11].

Questionnaire-measured 24 h-MBs

The DABQ consists of 14 basic items accompanied with different sub-items depending on the adults' employment status (employed or unemployed), and type of work schedule (regular working hours or irregular working hours) [7]. Questions assessed sleep and domain specific (occupational, commuting and other non-occupational) SB and PA of the past seven days. Reponses included time points (hh: mm), visual analogue scales, yes/no questions and numbers (i.e. number of days, hours, minutes) [7]. This questionnaire was tested for reliability and results showed moderate absolute agreement for test-retest reliability ranging from 0.59 to 0.69 [7]. The original language of this questionnaire is English and has therefore been translated to Dutch via back-to-back translation. For this study, the primary outcomes from the DABQ were used for interpretation including average minutes of daily activity behaviors (i.e. sleep, SB, LPA, MVPA in min/day) [Additional File 1] [7].

General information questionnaire

Sociodemographic information was asked via a short general questionnaire to determine the participants' characteristics: sex, age, body mass index (BMI), educational level (i.e. low (until secondary education), mid (college), high (university)), smoking status (i.e. non-smoker, ex-smoker, smoker), family situation (i.e. living alone, not having a partner; living together with partner; living alone, having a partner), children living with you (i.e. yes or no), neighborhood (i.e. rural, urban), sedentary job (i.e. no job, mostly spending time sitting at work, combining sitting and LPA at work, physically active job), net family income (i.e. <2000 euro, ≥2000 euro). Moreover, measurement characteristics were collected by asking the feasibility (very feasible until not feasible at all), annoyingness (very annoying until not annoying at all) and consistency (very consistent; more or less consistent; not consistent) of wearing the accelerometer during the past 7 days and nights.

Data processing

Accelerometer data processing

Raw accelerometer data were processed by the R package GGIR version 3.0–9 [22]. As raw accelerometer data can be analyzed using different data processing methods (cut-points to classify activity intensities and data reduction method-specific), the decision was made to process the data by four commonly used data processing methods validated for hip-worn accelerometry [19]. Therefore, the analysis process was identically repeated four times for each data reduction method (1) ENMO, (2) MAD, (3) CPM VA, (4) CPM VM [22, 23]. For each data reduction method, commonly used cut-points were chosen [19]. The ENMO metric used cut-points of Hildebrand et al. (2014) (2017), i.e. LPA (LPA) (47m*g*), MPA (MPA) (69m*g*) and VPA (VPA) (260 m*g*). The MAD metric used cut-points of Vähä Ypyä et al. (2018) (2023), i.e. LPA (22.5 m*g*), MPA (94 (m*g*), and VPA (396 m*g*). The CPM VA metric used cut-points of Troiano et al. (2008), i.e. LPA (100 CPM), MPA (2020 CPM), VPA (5999 CPM). The CPM VM metric used Sasaki et al. (2011) cut-points, i.e. LPA (200 CPM), MPA (2690 CPM) and VPA (6166 CPM) [13–18].

Accelerometer data were considered valid when having a minimum wear time of 16 h per day, for at least 4 days (including three weekdays and one weekend day) [24]. Non-wear time was defined as intervals of 60 min or more where the acceleration recorded as less than 13 m*g* on at least two out of three axes, or when the cumulative acceleration range falls below 50 m*g* [25]. Any periods noted in the diary where the device was removed were also classified as non-wear time [11]. Furthermore, data quality reports excluded ActiGraph files with a post-calibration error exceeding 0.01 g ($n=1$).

Sleep duration was determined using the R package GGIR, in combination with the reported wake-up and bedtime entries in the sleep diary. The HDCZA algorithm is used within this given window to detect sleep interruptions. If sleep diary data were invalid, the HDCZA algorithm was employed to detect the total sleep period based on the absence of changes in the z-angle greater than 5 degrees for at least 5 min compared to individualized thresholds (account for between-individual differences in z-angle distribution) [21].

DABQ data processing

The data of the DABQ were analyzed using the additional data processing Excel (DABQ analyzer 3.0) sheet created by the developers of the questionnaire [7]. After entering the values in the DABQ analyzer, an automatic data validation of correct values was performed and primary outcomes (i.e. daily average minutes of sleep, SB, LPA, MVPA) were calculated by DABQ analyze specific algorithms.

Statistics

Participant characteristics and movement behaviors were described using means and standard deviations for continuous data. Categorical variables were described using counts and percentages.

Absolute agreement and concurrent validity between measurement methods were explored for each individual behavior. The intraclass correlation (ICC) estimates and their corresponding 95% confidence intervals (95% CI) were calculated to estimate the absolute agreement between two measurement methods. This approach enables a comparison between DABQ use of time and the time measured by the ActiGraph for each behavior individually, using a two-way random effect model to assess absolute agreement [26]. Classification of ICC resembles < 0.50 as poor, between 0.50 and 0.75 as moderate and above 0.75 as good [26]. Bland-Altman plots were employed to visualize the mean differences (MD) and lower and upper limit of agreement (LOA) between the two methods. Spearman's rho (i.e. not normally distributed data) were used to identify the interclass correlation between measurements which focuses on the trend in the relationship, irrespective of agreement. Correlation cut offs were < 0.30 for poor correlation, between 0.30 and 0.60 for moderate correlation and between 0.61 and 0.85 for moderately high correlation and above 0.85 as high correlation [27]. Last, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were calculated. The MAE quantifies the magnitude of error without considering their direction. A smaller MAE indicate that the measurement closely approximated the reference measurement (ActiGraph). RMSE highlights whether the measurement under comparison (DABQ) challenged by significant outliers or large errors. A lower RMSE indicates that the measurement (DABQ) is consistently close to the reference measurement (ActiGraph).

Last, potential interaction effects between sociodemographic characteristics (i.e. sex, age, educational level, BMI, smoking status [see Additional file 2]) and the measurement method (DABQ vs. ActiGraph) were explored for each behavior separately by using mixed models with random intercepts at the participant level. If this interaction effect was statistically significant with a sociodemographic variable, validity statistics were explored for the different subgroups of this sociodemographic variable. For example, the sample was stratified into three subgroups for educational level (low, mid and high). For each of these subgroups, the ICC (95% CI), Spearman's rho, MD with upper and lower LOA, MAE and RMSE were reported.

All analyses were performed in R Statistical Software (v4.3.2) and repeated by the four different data processing methods (ENMO vs. MAD vs. CPM VA vs. CPM VM) [28].

Results

A total of 105 adults were recruited with a mean age of 45 (±13) years old, 54% were females and had a self-reported mean BMI of 25 kg/m² (±4) (Table 1). Participants described wearing the accelerometer during day and night as consistent, feasible and not annoying (Table 1). One adult was excluded from further analysis due reporting inconsistent wearing of the accelerometer during both day and night.

Regarding absolute agreement, a moderate ICC of 0.56 (0.41;0.67) was found for sleep across the four different data processing methods. Across all data processing methods, poor absolute agreement was found for SB, LPA and MVPA (Table 2). For SB, the ICC ranged between 0.01 (-0.04;0.07) and 0.38 (0.21;0.53) with ENMO predicting the lowest ICC and CPM VM the highest ICC. Regarding LPA, similar patterns were found, the lowest ICC of 0.00 (-0.03;0.05) for ENMO and the highest ICC of 0.31 (0.12;0.47) for CPM VM. Regarding MVPA, ranges between ICCs were smaller, with ENMO having an ICC of 0.23 (0.05;0.40), MAD an ICC of 0.24 (-0.04;0.47), CPM VA an ICC of 0.30 (0.12;0.47) and CPM VM an ICC of 0.26 (-0.05;0.50).

Spearman's correlations showed comparable findings for the interclass validity between DABQ- and ActiGraph-measured movement behaviors. A moderate interclass correlation was found for sleep (rho_{sleep}=0.58) across different data processing methods. For SB, LPA and MVPA the interclass correlation was poor for the ENMO (rho_{SB} = 0.01; rho_{LPA} = 0.10, rho_{MVPA} : 0.42), MAD (rho_{SB}: 0.39, rho_{LPA}: 0.46; rho_{MVPA}: 0.44), CPM VA (rho_{SB} :0.36; rho_{LPA}: 0.42; rho_{MVPA}: 0.43) and CPM VM (rho_{SB}: 0.43; rho_{LPA}: 0.42; rho_{MVPA}: 0.38) (Table 2).

Bland-Altman plots in Fig. 1 accompanied by the numbers in Table 2 show the MD and upper and lower LOA of movement behaviours between measurement methods and stratified by the four different data processing methods. Differences in sleep estimated by the DABQ-measurement compared to ActiGraph-measurement were small (mean difference: 2.18 min/day (±47.90)) across data processing methods. The MAE and RMSE were 35.26 and 47.72 for sleep respectively. Time spent in SB by the DABQ-measurement was lower compared to the ActiGraph-measurement across different data processing methods (ENMO −368.68 (±204.37) min/day; MAD −120.97 (±179.62) min/day; CPM VA -93.45 (±179.62) min/day; CPM VM -12.88 (±173.54) min/day). The MAE ranged between 143.41 and 368.69 and the RMSE ranged between 173.19 and 421.06 across data processing methods. Time spent in LPA was higher when measured with the DABQ-measurement compared to the ActiGraph-measurement across different data processing methods (ENMO + 401.17 (±201.06) min/day; MAD + 177.61 (±179.33) min/day; CPM VA + 122.63 (±180.75) min/day; CPM VM + 70.14 (±178.17) min/day). The MAE ranged between 151.76 and 401.17 and the RMSE ranged between 190.69 and 448.31 across data processing methods. Last, time spent in MVPA was lower by the DABQ-measurement by 10 until 40 min/day compared to the ActiGraph-measurement (ENMO −13.35 (±39.29) min/day; MAD −37.49 (±39.47) min/day; CPM VA -10.04 (±36.68) min/day; CPM VM -38.11 (±37.87) min/day) (MAE range: 26.34–46.77; RSME range: 41.32–54.29 across data processing methods).

Significant interactions between measurement method and participant characteristics were found for

Table 1 Sociodemographic and measurement characteristics of the total sample ($n=105$)

	***N*=105**
Sociodemographic characteristics	
Age (mean, SD)	45.38 (12.9)
Sex=female (n, %)	57 (54.3)
BMI (kg/m²) (mean, SD)	25.03 (4.3)
Missings (n,%)	*2 (1.9)*
Family situation (n,%)	
Living alone, not having partner	12 (11.4)
Living together with partner	89 (84.8)
Living alone, having a partner	4 (3.8)
Living with children=yes (n,%)	83 (79.0)
Neighborhood (n,%)	
Rural	82 (78.1)
Urban	23 (21.9)
Educational level (n, %)	
Low	27 (25.7)
Mid	37 (35.2)
High	41 (39.0)
Sedentary job (n, %)	
No job	11 (10.5)
Mostly spending time sitting at work	51 (48.6)
Combining sitting and LPA at work	31 (29.5)
Physically active job	12 (11.4)
Net family income=more then 2000 euro (n, %)	84 (80.0)
Missings (n,%)	*21 (20.0)*
Smoking (n, %)	
No	89 (84.8)
Ex-smoker	11 (10.5)
Yes	5 (4.8)
Measurement characteristics	
Feasibility to wear the accelerometer at the hip during waking hours for 7 days (n, %)	1.6 (0.8)
Very feasible	59 (56.2)
Feasible	37 (35.2)
Neutral	6 (5.7)
Not feasible	2 (1.9)
Not feasible at all	1 (1.0)
Feasibility to wear the accelerometer at the wrist during the night for 7 days (n, %)	
Very feasible	61 (58.1)
Feasible	29 (27.6)
Neutral	14 (13.3)
Not feasible	0 (0.0)
Not feasible at all	1 (1.0)
Annoying to wear the accelerometer at the hip during waking hours for 7 days (n, %)	3.96 (1.0)
Not annoying at all	38 (36.6)
Not annoying	37 (35.6)
Neutral	19 (18.3)
Annoying	7 (6.7)
Very annoying	3 (2.9)
Annoying to wear the accelerometer at the wrist during the night for 7 days (n, %)	
Not annoying at all	42 (40.4)
Not annoying	39 (37.5)
Neutral	17 (16.3)
Annoying	4 (3.8)
Very annoying	2 (1.9)
Perceived consistency of wearing the accelerometer during the day (n, %)	

Table 1 (continued)

	N = 105
Very consistent	100 (95.2)
More or less consistent	4 (3.8)
Not consistent	1 (1.0)
Perceived consistency of wearing the accelerometer at night (n, %)	
Very consistent	98 (93.3)
More or less consistent	6 (5.7)
Not consistent	1 (1.0)

educational level ($Sleep_{ENMO,\ MAD,\ CPMVA,\ CPMVM}$ est. 26.21 (1.75; 50.66); SB_{ENMO} est. 156.31 (59.14; 253.48); LPA_{ENMO} est. -159.03 (-254.36;-63.71); SB_{MAD} est. 103.49 (16.56; 190.42); LPA_{MAD} est. -104.46 (-191.42;-17.49); SB_{CPMVA} est. 121.16 (35.35; 206.96); LPA_{CPMVA} est. -128.29 (-214.37;-42.22); SB_{CPMVM} est. 108.69 (25.38; 191.98); LPA_{CPMVM} est. -116.09 (-201.72; -30.46)) across data processing methods [See Additional file 2].

Regarding sleep, individuals with higher educational levels most accurately reported sleep duration. Those with mid-level educational levels overreported and those with low educational levels underreported sleep duration by the DABQ compared to the ActiGraph. SB was consistently underreported and LPA overreported by the DABQ. However, the magnitude of under- and overreporting decreased from low educational level to mid educational to high educational level. This was seen across all data processing methods, except for CPM VM where the high educational level group overreported SB.

Despite the differences in behaviors, the overall absolute agreement and validity was moderate for sleep and poor for the other behaviors across the different educational subgroups. See Additional File 3 for all validity statistics across the educational subgroups.

Discussion

This study found a moderate absolute agreement and concurrent validity for sleep comparing the DABQ with ActiGraph. Conversely, a poor absolute agreement and validity was found for SB, LPA, and MVPA. Specifically, the DABQ tended to underestimate SB and MVPA while overestimating LPA. The extent of these discrepancies varied depending on the accelerometer data processing method and the cut-points used. For example, the underestimation of SB by the DABQ, as compared to the ActiGraph, ranged significantly: 368 (±204) min/day using ENMO, 120 (±179) min/day using MAD, 93 (±179) min/day using CPM VA, and 12 (±173) min/day using CPM VM. Regarding sleep, the MAE and RMSE showed rather large values compared to MD. As MAE and RMSE quantifies the magnitude of error without considering the direction, we can say that despite the small MD in sleep, their error measurements were higher. This suggests that when comparing sleep within large groups, the questionnaire might be acceptable to use because the bias (mean difference) is small. However, for sleep measurements at precise and individual level, the questionnaire is not recommended due to high variability and error rate. Regarding the other behaviors, both MD, MAE, and RMSE were high resulting in less acceptability to use the questionnaire in large groups and at individual level. Last, self-reported sleep, SB and LPA varied by educational level. Higher educational levels was associated with more accurate sleep reporting, while SB was consistently underreported and LPA consistently overreported, across educational levels. Importantly, the magnitude of these discrepancies diminished as educational levels increased, demonstrating a potential link between education and reporting accuracy. However, despite these trends, the overall validity of self-reports was only moderate for sleep and poor for SB and LPA across all educational subgroups.

In general, our findings of poor absolute agreement between objective and subjective measurements align with existing literature. A systematic review comparing accelerometry and PA questionnaires reported weak to moderate correlations (r=0.14; r=0.58) between these measurement methods, but with no clarification on the extent of over- or underreporting of PA by self-reported questionnaires [12]. Other studies have found similar low to moderate correlations for time spent in SB (r=0.32±0.20; [5] r=0.35, [95% CI 0.32–0.39] [29]). Similar to our results, Prince et al. (2020) reported an underestimation of SB by approx. 2 h/day in self-reported measurements compared to device-based measurements [5]. For sleep, correlations between self-reported sleep duration and accelerometry have been reported with ranges from poor to moderately correlated (r=0.45 [6], r=0.14 [30], r=0.40 [31]), and poor correlated with polysomnography (r=0.20 [6]). Studies have found that self-reports typically overestimate sleep duration by about one hour [6, 30, 31]. This could be explained by the difficulties in accurately capturing sleep latency and wake time after sleep onset by individuals [6]. Furthermore, the discrepancies in reporting time spent in 24 h-MBs may also stem from recall bias and social desirability [5, 12]. Thus, some researchers recommend using self-reported questionnaires to rank activity intensities (e.g. in quartiles

Table 2 Validity statistics comparing DABQ with actigraph regarding four data processing methods (ENMO, MAD, CPM VA, CPM VM) ($n = 101$)

24 h- MBs	ActiGraph (SD), min/day	DABQ (SD), min/day	ICC (95% CI)	Spearman's Rho	MD (SD)	Upper LOA	Lower LOA	MAE	RMSE
ENMO									
Sleep	467.18 (50.90)	465.01 (50.36)	0.56 (0.41;0.67)	0.58	2.18 (47.90)	-91.7	96.06	35.26	47.72
SB	883.36 (67.02)	514.68 (197.21)	0.01 (-0.04;0.07)	0.01	368.68 (204.37)	-31.9	769.25	368.69	421.06
LPA	34.25 (16.04)	435.43 (202.12)	0.00 (-0.03;0.05)	0.1	-401.17 (201.06)	-795.26	-7.09	401.17	448.31
MVPA	38.23 (23.23)	24.88 (38.78)	0.23 (0.05; 0.40)	0.42	13.35 (39.29)	-63.66	90.36	29.18	41.32
MAD									
Sleep	467.18 (50.90)	465.01 (50.36)	0.56 (0.41;0.67)	0.58	2.18 (47.90)	-91.7	96.06	35.26	47.72
SB	635.66 (90.93)	514.68 (197.21)	0.24 (0.02;0.44)	0.39	120.97 (179.62)	-231.08	473.03	167.58	215.84
LPA	257.82 (66.56)	435.43 (202.12)	0.17 (-0.05;0.38)	0.46	-177.61 (179.33)	-529.10	173.88	202.79	251.78
MVPA	62.37 (31.22)	24.88 (38.78)	0.24 (-0.04; 0.47)	0.44	37.49 (39.47)	-39.86	114.84	46.77	54.29
CPM VA									
Sleep	467.18 (50.90)	465.01 (50.36)	0.56 (0.41;0.67)	0.58	2.18 (47.90)	-91.7	96.06	35.26	47.72
SB	608.13 (88.15)	514.68 (197.21)	0.26 (0.06;0.44)	0.36	93.45 (179.62)	-258.60	445.49	159.67	201.71
LPA	312.80 (71.36)	435.43 (202.12)	0.22 (0.01; 0.41)	0.42	-122.63 (180.75)	-476.90	231.63	173.31	217.71
MVPA	34.92 (19.02)	24.88 (38.78)	0.30 (0.12; 0.47)	0.43	10.04 (35.68)	-59.90	79.97	26.34	36.90
CPM VM									
Sleep	467.18 (50.90)	465.01 (50.36)	0.56 (0.41;0.67)	0.58	2.18 (47.90)	-91.7	96.06	35.26	47.72
SB	527.56 (98.59)	514.68 (197.21)	0.38 (0.21;0.53)	0.43	12.88 (173.54)	-327.26	353.02	143.41	173.19
LPA	365.29 (84.91)	435.43 (202.12)	0.31 (0.12;0.47)	0.42	-70.14 (178.17)	-419.36	279.08	151.76	190.69
MVPA	62.99 (30.47)	24.88 (38.78)	0.26 (-0.05;0.50)	0.38	38.11 (37.87)	-36.13	112.35	45.84	53.60

SB: sedentary behavior, LPA: light physical activity, MVPA: moderate to vigorous physical activity, ICC: intraclass correlation, 95% CI: 95% Confidence intervals, MD: mean difference between two measurement methods, SD: standard deviation of the mean difference, LOA: limit of agreement. MAE: Mean Absolute Error; RMSE: Root Mean Square Error. The intensity-based cutoff points thresholds for each metric are as follows: ENMO Hildebrand et al. (2014), MAD Vähä Ypyä et al. (2018, 2023), CPM VA Troiano et al. (2008), and CPM VM Sasaki et al. (2011)

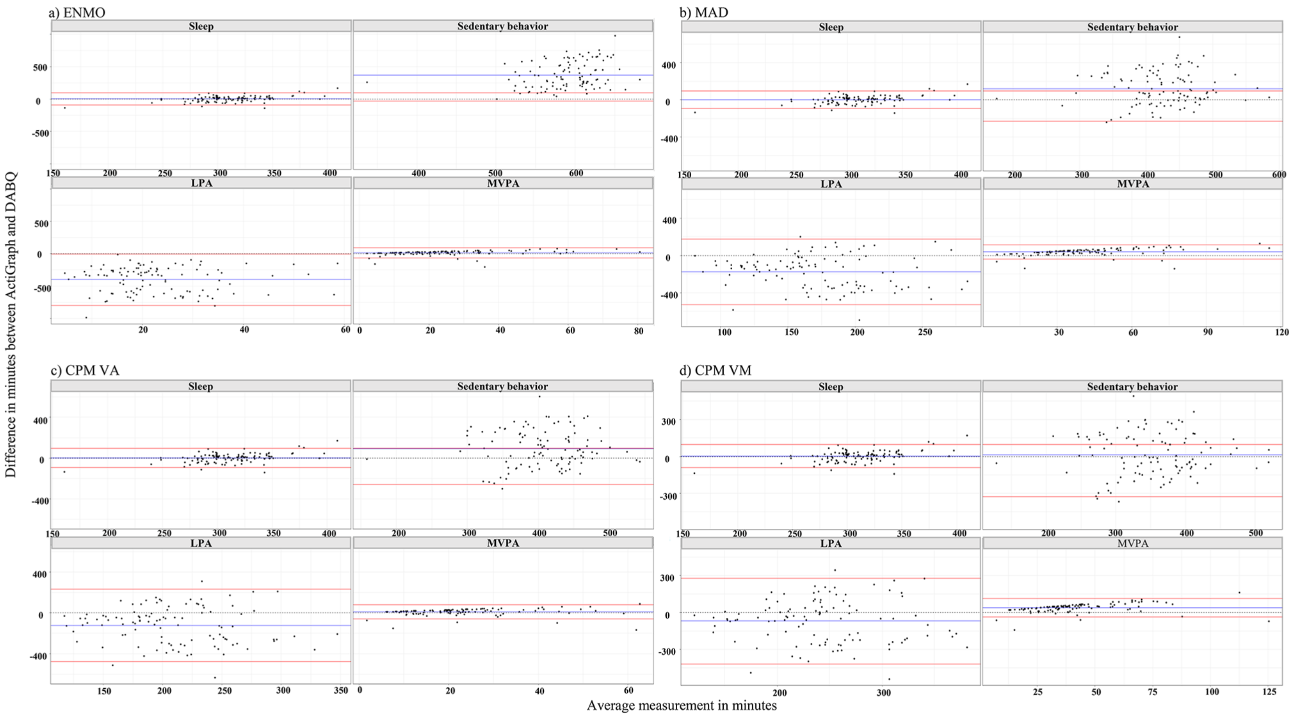

Fig. 1 Bland-Altman plots showing difference in minute between ActiGraph- and DABQ-measured movement behaviors regarding different data processig methods (**a**) ENMO, (**b**) MAD, (**c**) CPM VA, (**d**) CPM VM. X-axis refers to average measurement of time use estimates measured by ActiGraph and DABQ. Y-axis shows the differences in time-use estimates between the ActiGraph and DABQ. The Intensity-based cut-points thresholds for each data processing method are as follows: ENMO Hildebrand et al. (2014), MAD Vaha Ypya et al. (2018, 2023), CPM VA Troiano et al. (2008), and CPM VM Sasaki et al. (2011). Blue line represent the mean difference (MD) and the red lines represent upper limit of agreement, and lower limit of agreement. For exact MD and upper and lower LOA see Table 2

of activity level) rather than attempting to recall absolute time periods (min/day) as this can minimize this impact of inaccuracies of self-reporting [5, 32].

One paper by Kastelic et al. (2022) compared DABQ measurements to those obtained from an ActivPAL device and found similar correlation coefficients and ICCs to our study [7]. However, when comparing the average time spent in movement behaviors as reported by the DABQ and ActiGraph, it is important to carefully examine the breakdown of the 24 h-MBs resulted from the different accelerometer data processing methods, especially when using the ENMO. Notably, the data processing method ENMO, using the cut-points established by Hildebrand et al. (2014, 2017), indicated an average of 883 min/day spent in SB and only 34 min/day in LPA, which appears unrealistic [14]. This time spent in SB and LPA using ENMO with the cut-point of Hildebrand et al. (2014, 2017), constrasts significantly with the findings of Kastelic et al. (2022), who used the ActivPAL for comparison [7, 10]. Since the ActivPAL is widely regarded as the preferred method for SB classification, this highlights the importance of critically evaluating the measurement methodologies and data processing methods employed to process raw accelerometry to ensure the accurate assessment of true activity levels.

It is important to emphasize the impact of measuring time spent in various behaviors using tools such as the DABQ or accelerometers on factors like guideline compliance and their associations with health outcomes. Variations in time spent in 24 h-MBs can significantly affect research findings. For example, a previous study examined the effects of using different data processing methods (e.g. ENMO, MAD, CPM VA, CPM VM) on guideline compliance and cardiometabolic health associations [19]. The study found that the choice of data processing method led to differing time allocations within the 24 h-MBs, resulting in varying guideline compliance rates ranging from 0 to 25% [19]. Furthermore, differences in associations with cardiometabolic indicators like BMI and WC were observed based on the method used [19]. However, these results only apply to research using the ActiGraph and do not apply for the DABQ. But, based on these results, it is hypothesized that the choice of method could substantially influence the observed differences in associations with health outcomes.

Last, we found that educational level moderated the variations in 24 h-MBs between measurement methods. This is complementary with previous research which has identified participant characteristics like age, sex, obesity status, and depression that drive differences in the direction of correlations and measurement accuracy [5, 12, 20, 31, 33]. In our study, the magnitude of over- and underreporting varied according to educational level with a decreasing trend as the educational level increases. These findings suggest that participant characteristics might significantly impact the outcomes. When using the DABQ-measurement as a single measurement method of 24-h MB, results must be interpreted with caution considering possible over- and underestimations, especially in heterogenous study populations (e.g. wide variability in educational level).

Strengths and limitations

This is the first study to assess the validity of the DABQ compared with the ActiGraph. The ActiGraph has been processed by four different data processing methods (cut-points and data reduction methods specific) which is consistently analyzed using GGIR [22]. A significant limitation was that most participants did not report the reason for their non-wear time of the device in the diary. This made it difficult to classify non-wear time as sedentary time or water-based activities (e.g. swimming) as an addition to the device-based non-wear time measurement. However, GGIR applied the same algorithm across the different data processing methods to define non-wear time, which makes non-wear time classification consistent across ENMO, MAD, CPM VA and CPM VM. Next, since there is no gold standard for cut-points to classify activity intensities, determining the extent of deviations between both measurement methods across the different data processing methods posed a challenge. However, as seen in other research, this paper also shows the least realistic SB and LPA estimates when processing hip-worn ActiGraph data with the ENMO data reduction method and Hildebrand et al. (2014, 2017) cut-points [19]. Last, the secondary outcomes of the DABQ including sleep indicator and time spent sedentary and physically active across different domains, i.e. occupational time, commuting time, non-occupational activities (e.g., walking sport, other physically demanding activities, recreational screen time, muscle strengthening exercise) could provide additional contextual information [See Additional File 1] [5, 7, 33]. However, based on the results of this current study, it is not possible to know in which domain the over- or underreporting of time spent in behaviors takes place. Therefore, further research is recommended to explore the capability of this questionnaire to accurately assess all aspects (i.e. domain, context, duration, frequency and volume) of the behaviors in one day.

Conclusion

Compared to the ActiGraph, the DABQ showed moderate absolute agreement and concurrent validity regarding sleep, yielding accurate time-use estimates within large groups. However, the DABQ tended to underestimate SB and MVPA while overestimating LPA, which was reflected in poor absolute agreement and concurrent validity for these measures. Additionally, differences in time-use estimates between measurement methods varied by educational level and the data processing method. A higher education level was linked with more accurate sleep reporting, but with SB consistently underreported and LPA overreported across all educational levels. Discrepancies in under- and overreporting decreased with increasing educational levels, suggesting a potential connection between education and reporting accuracy. Despite this, self-report validity remained moderate for sleep and poor for SB and LPA across all educational subgroups. Therefore, using DABQ as a single measurement of the 24 h-MBs should be interpreted with caution. Nevertheless, each measurement method has its unique advantages and disadvantages which should be considered when selecting the most appropriate tool to address a study's aims. More research is needed to finetune the questionnaire to eventually harmonize self-reported and objectively measured methods to capture 24 h-MBs.

Abbreviations

PA	Physical activity
SB	Sedentary behavior
24 h-MBs	24-hour movement behaviors
MVPA	Moderate to vigorous physical activity
LPA	Light physical activity
ENMO	Euclidian norm minus one
MAD	Mean amplitude deviation
CPM	Countute
VA	Vertical axis
VM	Vector magnitude
BMI	Body mass index
WC	Waist circumference
DABQ	Daily activity behavior questionnaire
ICC	Intraclass correlation coefficient
CI	Confidence interval
MD	Mean difference
LOA	Limit of agreement
MAE	Mean absolute error
RMSE	Root mean squared error
min/day	Minutes a day
h/day	Hours a day

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-21139-8.

Supplementary Material 1: **Additional File 1**: Secondary outcomes of the DABQ

Supplementary Material 2: **Additional File 2**: Interaction effects of measurement method (ActiGraph versus DABQ) and participants characteristics regarding different data processing methods a) ENMO b) MAD c) CPM VA d) CPM VM (n = 101)

Supplementary Material 3: **Additional File 3**: Validity statistics comparing DABQ with ActiGraph for different educational subgroups (n = 101)

Acknowledgements
We would like to thank all participants included in this study.

Author contributions
IW conceptualized the idea for this manuscript, collected and cleaned the data, performed the data analysis, interpreted the data and drafted the manuscript; MDC provided in-depth guidance during the previously mentioned processes. All the authors critically read, provided revisions and approved the final submitted version of the manuscript (VV, DD, PC, BL, MDC).

Funding
I.W. is supported by the Research Foundation Flanders (FWO-11N0422N). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability
The dataset and configuration file are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was included in the approval of the Ethical Review Committee of Ghent University, Belgium, in line with national regulations (the Ethical Committee of Ghent University Hospital (Belgium), ONZ-2023-0384). Informed consent was provided, explained and signed by all participants prior to the start of the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium
[2]Research Foundation Flanders, Brussels, Belgium
[3]Department of Health Promotion, Maastricht University, Research Institute of Nutrition and Translation Research in Metabolism (NUTRIM), Maastricht, The Netherlands
[4]Department of Health Promotion, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands
[5]Alliance for Research in Exercise, Nutrition and Activity, Allied Health & Human Performance, University of South Australia, Adelaide, SA, Australia
[6]Department of Internal Medicine and Pediatrics, Department of Endocrinology, Ghent University Hospital & Ghent University, Ghent, Belgium

Received: 28 June 2024 / Accepted: 18 December 2024
Published online: 30 December 2024

References

1. Rollo S, Lang JJ, Roberts KC, Bang F, Carson V, Chaput JP, et al. Health associations with meeting the Canadian 24-hour movement guidelines for adults: results from the Canadian Health measures Survey. Health Rep. 2022;33(1):16–26.
2. Walia HK, Mehra R. Practical aspects of actigraphy and approaches in clinical and research domains. Handb Clin Neurol. 2019;160:371–9.
3. Bakhtari Aghdam F, Aziz-Zadeh S, Musavi S, Abbasalizad-Farhangi M. Psychometric properties of the last-7-Day sedentary time questionnaire (SIT-Q-7d): testing the validity and reliability among general population. BMC Public Health. 2022;22(1):1880.
4. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act. 2011;8:115.
5. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2020;17(1):31.
6. Jackson CL, Patel SR, Jackson WB 2nd, Lutsey PL, Redline S. Agreement between self-reported and objectively measured sleep duration among white, black, hispanic, and Chinese adults in the United States: multi-ethnic study of atherosclerosis. Sleep. 2018;41(6).
7. Kastelic K, Šarabon N, Burnard MD, Pedišić Ž. Validity and reliability of the Daily Activity Behaviours Questionnaire (DABQ) for Assessment of Time Spent in Sleep, sedentary Behaviour, and physical activity. Int J Environ Res Public Health. 2022;19(9).
8. An HS, Kim Y, Lee JM. Accuracy of inclinometer functions of the activPAL and ActiGraph GT3X+: a focus on physical activity. Gait Posture. 2017;51:174–80.
9. Radtke T, Rodriguez M, Braun J, Dressel H. Criterion validity of the ActiGraph and activPAL in classifying posture and motion in office-based workers: a cross-sectional laboratory study. PLoS ONE. 2021;16(6):e0252659.
10. Chastin SFM, Dontje ML, Skelton DA, Čukić I, Shaw RJ, Gill JMR, et al. Systematic comparative validation of self-report measures of sedentary time against an objective measure of postural sitting (activPAL). Int J Behav Nutr Phys Act. 2018;15(1):21.
11. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer Data Collection and Processing Criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med. 2017;47(9):1821–45.
12. Skender S, Ose J, Chang-Claude J, Paskow M, Brühmann B, Siegel EM, et al. Accelerometry and physical activity questionnaires - a systematic review. BMC Public Health. 2016;16(1):515.
13. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. Scand J Med Sci Sports. 2017;27(12):1814–23.
14. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc. 2014;46(9):1816–24.
15. Vähä-Ypyä H, Bretterhofer J, Husu P, Windhaber J, Vasankari T, Titze S et al. Performance of different Accelerometry-Based Metrics to Estimate Oxygen Consumption during Track and Treadmill Locomotion over a wide intensity range. Sens (Basel). 2023;23(11).
16. Vähä-Ypyä H, Husu P, Suni J, Vasankari T, Sievänen H. Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer. Scand J Med Sci Sports. 2018;28(3):1092–102.
17. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008;40(1):181–8.
18. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. 2011;14(5):411–6.
19. Willems I, Verbestel V, Dumuid D, Calders P, Lapauw B, De Craemer M. A comparative analysis of 24-hour movement behaviors features using different accelerometer metrics in adults: implications for guideline compliance and associations with cardiometabolic health. PLoS ONE. 2024;19(9):e0309931.
20. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol. 2014;179(6):781–90.
21. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A Novel, Open Access Method to Assess Sleep Duration using a wrist-worn accelerometer. PLoS ONE. 2015;10(11):e0142533.
22. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep outcomes from Multi-day Raw Accelerometer Data. J Meas Phys Behav. 2019;2(3):188–96.
23. Neishabouri A, Nguyen J, Samuelsson J, Guthrie T, Biggs M, Wyatt J, et al. Quantification of acceleration as activity counts in ActiGraph wearable. Sci Rep. 2022;12(1):11958.
24. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large Scale Population Assessment of Physical Activity using wrist worn accelerometers: the UK Biobank Study. PLoS ONE. 2017;12(2):e0169649.
25. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol (1985). 2014;117(7):738–44.
26. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.
27. Portney LG, Watkins MP. Foundations of Clinical Research: applications to practice. Pearson/Prentice Hall; 2009.

28. Team RC. R: a Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
29. Bakker EA, Hartman YAW, Hopman MTE, Hopkins ND, Graves LEF, Dunstan DW, et al. Validity and reliability of subjective methods to assess sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2020;17(1):75.
30. Lee PH. Validation of the National Health and Nutritional Survey (NHANES) single-item self-reported sleep duration against wrist-worn accelerometer. Sleep Breath. 2022;26(4):2069–75.
31. Matthews KA, Patel SR, Pantesco EJ, Buysse DJ, Kamarck TW, Lee L, et al. Similarities and differences in estimates of sleep duration by polysomnography, actigraphy, diary, and self-reported habitual sleep in a community sample. Sleep Health. 2018;4(1):96–103.
32. Sagelv EH, Hopstock LA, Johansson J, Hansen BH, Brage S, Horsch A, et al. Criterion validity of two physical activity and one sedentary time questionnaire against accelerometry in a large cohort of adults and older adults. BMJ Open Sport Exerc Med. 2020;6(1):e000661.
33. Colley RC, Butler G, Garriguet D, Prince SA, Roberts KC. Comparison of self-reported and accelerometer-measured physical activity in Canadian adults. Health Rep. 2018;29(12):3–15.

Publisher's note