Psychiatry Research 271 (2019) 311–318

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Revisiting the International Physical Activity Questionnaire (IPAQ): Assessing sitting time among individuals with schizophrenia

Markus J. Duncan[a,*], Kelly Arbour-Nicitopoulos[b], Mehala Subramaniapillai[c], Gary Remington[d,e], Guy Faulkner[a]

[a] School of Kinesiology, University of British Columbia, 2148 Health Sciences Mall, Room 4008, Vancouver, BC V6T 1Z8, Canada
[b] Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON, Canada
[c] Mood Disorders Psychopharmacology Unit, University Health Network, University of Toronto, Toronto, Canada
[d] Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[e] Schizophrenia Division, Centre for Addiction and Mental Health, Toronto, ON, Canada

ARTICLE INFO

Key Words:
Schizophrenia
Physical activity
Sedentary behaviour
Sitting
Validation
Measurement

ABSTRACT

While moderate to vigorous physical activity may be one method of addressing common physical morbidities in schizophrenia, reducing sedentary time may be a low intensity adjunct. In order to determine whether sedentary behaviour is associated with health outcomes, valid and reliable tools for assessing sedentary time are necessary. In order to characterize the validity and reliability of the International Physical Activity Questionnaire (IPAQ) for assessing sitting (sedentary) time, participants completed the IPAQ at baseline and 4 weeks later and wore accelerometers for 7 days before the final assessment. Bland-Altman analyses and intraclass correlation coefficients (ICC) were used to compare agreement between measurements. One-hundred thirteen individuals completed the study. Mean difference between the IPAQ and accelerometer was 26.8 min (95% Limits of Agreement: −458.7–512.3) and $ICC_{A,1}$ was 0.23 (95% CI: 0.06–0.39). Week 1 and Week 4 administrations of the IPAQ differed by an average of 26.6 min, (95% Limits of Agreement: −510.9–564.2) and $ICC_{A,1}$ was 0.41 (95% CI: 0.21–0.59). The "minutes" of sitting reported by the IPAQ do not reflect objective sedentary behaviour measurements and this current measure may be unsuitable for the population level assessment of sitting time among individuals with schizophrenia.

1. Introduction

Individuals with schizophrenia suffer from higher rates of obesity, diabetes, and cardiovascular disease compared to the general population (Correll et al., 2017; Dixon et al., 2000; Hennekens et al., 2005; Manu et al., 2015; Vancampfort et al., 2016a, 2015) While moderate-to-vigorous physical activity (PA) is an important component of health promotion and weight management, meta-analysis has demonstrated that greater amounts of daily sedentary behaviour (SB) – any waking activity characterized by an energy expenditure less than 1.5 metabolic equivalents in a sitting, reclining or lying posture (Tremblay et al., 2017) – are associated with higher rates of cardiovascular disease, diabetes, and all-cause mortality (Biswas et al., 2015; Warren et al., 2010). This association appears to be independent of moderate-to-vigorous PA (Biswas et al., 2015). Even individuals who are currently meeting physical activity guidelines may receive cardio-metabolic health benefits from interrupting sitting time with short breaks (Owen et al., 2010; Thorp et al., 2011). As such, studies have begun to demonstrate an association between schizophrenia and health outcomes such as body mass index among individuals with schizophrenia as well (Bueno-Antequera et al., 2017; Vancampfort et al., 2012, 2014). Therefore, while continuing to test interventions to increase moderate to vigorous PA among individuals with schizophrenia is still important (McNamee et al., 2013; Vancampfort et al., 2016b), physically active individuals with schizophrenia may accrue additional health benefits by reducing prolonged sitting time. Furthermore, reducing SB among inactive individuals may be easier than promoting more vigorous PA while still accruing some health benefit.

A recent meta-analysis demonstrated that individuals with psychosis engage in high levels of SB (Stubbs et al., 2016). Across 13 studies the authors reported a pooled effect size of 11 h of SB per day in this population. Among the four studies that compared individuals with psychosis to healthy controls, individuals with psychosis were engaging in 2.8 h of more SB. These numbers increase to 12.6 h of SB and 2.9 h

* Corresponding author.
E-mail address: mark.duncan@alumni.ubc.ca (M.J. Duncan).

https://doi.org/10.1016/j.psychres.2018.11.063
Received 1 August 2018; Received in revised form 26 November 2018; Accepted 26 November 2018
Available online 27 November 2018
0165-1781/

more than healthy controls when only objective measures of SB are used. The authors suggested that self-report questionnaires may be largely underestimating the amount of SB that individuals with psychosis are engaging in, though no studies included in the review used both objective and self-report methods simultaneously to evaluate this hypothesis. A subsequent meta-analysis of SB in schizophrenia, bipolar disorder and major depressive disorder similarly found lower self-reported SB compared to when measured objectively (Vancampfort et al., 2017). While the objective measurement of SB through accelerometry is the gold standard for assessment in the field, self-report will likely continue to be used for pragmatic reasons, particularly in epidemiological research.

The most common self-report instrument in the studies summarized by Stubbs et al. (2016) was the single item sitting scale of the International Physical Activity Questionnaire Short Form (IPAQ-SF) (Craig et al., 2003). The question asks, "During the last 7 days, how much time did you spend sitting on a weekday?" specifying to "Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television" in the preamble. Reliability and agreement have been previously characterized between the IPAQ-SF and accelerometry among samples of individuals with schizophrenia (Duncan et al., 2017; Faulkner et al., 2006), but only in regard to the moderate and vigorous PA scales of the IPAQ-SF. Additionally, Firth and colleagues (Firth et al., 2018) have expanded on this work by demonstrating that while PA scales of the IPAQ correlate with accelerometry, these scales are not suitable for comparing PA behaviours between individuals with schizophrenia to controls without schizophrenia. However, to our knowledge, and despite its apparent widespread use, similar assessments of the IPAQ sitting scale among individuals with schizophrenia has yet to be reported.

Contrary to Stubbs et al. (2016) criticism of self-report questionnaires underestimating SB, existing agreement data in the general adult population between the IPAQ-SF sitting scale and accelerometer-defined SB ($\leq$99 counts/min) indicates that the IPAQ-SF may overestimate the amount of sedentary time by +130 min/d with large 95% limits of agreement (LoA) −275–536 min/d (Hagstromer et al., 2010). Additionally, bias appeared to be proportional, with greater overestimation associated with greater reported sitting time. A systematic review of validation and reliability studies with the IPAQ-SF found Spearman's correlations (ρ) of 0.07–0.61 with accelerometer-defined SB and test-retest reliability ranging from Spearman's ρ of 0.18–0.95, and intraclass correlation coefficients (ICC) ranging from 0.80–0.97 (Healy et al., 2011).

However, cognitive deficits among people with schizophrenia are common. These include memory impairment (Keefe and Fenton, 2007; Reichenberg et al., 2006) which may impact participant's ability to accurately recall activities over an extended period, and deficits in attention and executive function (Keefe and Fenton, 2007; Reichenberg et al., 2006) which may impact the reliability and validity of any self-report questionnaire. Given these concerns, performing additional assessments of the IPAQ-SF sitting scale is warranted, and will shed light on the validity of studies that seek to examine relationships between SB and other outcomes of interest among samples of individuals with schizophrenia. Furthermore, evaluating the sitting scale of the IPAQ-SF will help inform ongoing efforts to develop new self-report measures for use among individuals with serious mental illness such as those described by Rosenbaum and Ward (2016).

The IPAQ-SF is designed for population level assessment (Bauman et al., 2009a, 2009b; Craig et al., 2003) as "small sample studies may be under-powered to detect between-group differences" (Bauman et al., 2009a, p. S6) due to the high variance in scores obtained. However, small samples comparing the IPAQ to accelerometry will offer some insight on how an instrument may perform at a population level. Furthermore, "establishing the reliability and validity of extant measures" as suggested by phase 2 of Sallis and colleague's behavioural epidemiological framework (Sallis et al., 2000, p. 295) may demonstrate new utility for existing tools even if not their initial intention – but this must be evaluated.

As such, this study is a follow-up to the study characterizing accuracy of the PA scales of the IPAQ-SF presented by Duncan and colleagues (Duncan et al., 2017). In the current study, data from a pre-existing prospective study (Arbour-Nicitopoulos et al., 2017) was reanalyzed with the purpose of assessing: (1) the agreement between accelerometry and the IPAQ-SF sitting scale, (2) test-retest reliability of the IPAQ-SF sitting scale over a 4-week time period. Furthermore, with respect to the first objective, the IPAQ-SF sitting scale may be better at assessing prolonged periods of sitting rather than shorter incidental bouts that participants may not remember. Thus, accelerometry derived SB is evaluated using bout (distinct periods of uninterrupted sitting) lengths of (1) $\geq$1-min, i.e. any 60 s epoch classified SB, regardless of total bout length; (2) $\geq$10-min, which correspond to common guidelines for what qualifies as a bout of physical activity (e.g. Canadian Society of Exercise Physiology, 2012); and (3) $\geq$30-min, which represent longer periods of sitting that may be easier to recall than incidental amounts of sitting accrued in smaller bouts, and correspond to the length of a typical TV-show, an archetypal form of SB, with each of these accelerometer-derived bout lengths compared to the IPAQ-SF.

Based on our previous work with PA accuracy in this population (Duncan et al., 2017) and the criticism by Stubbs et al. (2016), it was hypothesized that the IPAQ-SF sitting scale would underestimate accelerometer derived SB on average, but as per Hagstromer et al. (2010), the bias would be proportional to the amount of SB engaged in by participants. Further to this point, agreement would be improved when accelerometry derived SB is defined by longer bouts over shorter bouts. Given the longer reliability period of the study, test-retest reliability would be similar to what was previously reported for MVPA in the same sample (i.e., Spearman's ρ of 0.47) (Duncan et al., 2017).

2. Methods

2.1. Participants

Research ethics boards at the Centre for Addiction and Mental Health in Toronto and the University of Toronto approved the larger 4-week prospective study (Arbour-Nicitopoulos et al., 2017), the purpose of which was to evaluate psychological determinants of physical activity participation in the sample – no effort was made to change behaviour. Participants were required to: (1) be age 18–64 years (in line with the Canadian Physical Activity Guidelines recommendations for adults (Canadian Society of Exercise Physiology, 2012), and (2) have a diagnosis of schizophrenia or schizoaffective disorder. Diagnosis was confirmed with the Mini-International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998). Participants were excluded for: (1) being hospitalized over the past 12 months for angina pectoris, myocardial infarction, or cardiac surgery of any kind; and/or (2) uncontrolled hypertension (defined as blood pressure > 140/90). If eligible, participants provided written consent prior to commencing the study and capacity to consent was assessed with the MacArthur Competence Assessment Tool for Clinical Research (MacCAT-CR; Appelbaum and Grisso, 2001). Participants were recruited by referrals from nurses, psychiatrists, and other studies involving persons with schizophrenia at the Centre for Addiction and Mental Health. Given that this is a secondary data analysis, achieved power ($1 - \beta$) was calculated post-hoc using G*Power 3 (Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany) with α set at 0.05.

2.2. Procedures

SB levels were assessed at baseline with the IPAQ-SF and again four weeks later. Participants were instructed to wear an ActiGraph

(Pensacola, USA) wGT3X + accelerometer for 7 days over their right hip three weeks after baseline assessment of the primary study. Proper wear was demonstrated to the participant when receiving the accelerometer. Participants were also instructed to keep a log tracking their accelerometer wear time. Participants returned the accelerometer when completing the final IPAQ-SF. Thus, accelerometry data coincided with the period of time assessed by the Week 4 IPAQ-SF, allowing for direct comparison of the measurements of SB obtained by each instrument.

2.3. Analysis

Analysis methods are similar to those described previously for evaluating the PA in this sample (Duncan et al., 2017): Accelerometry data was analyzed using ActiGraph's (Pensacola, USA) *Actilife* software (v6.12). Actilife's sleep analysis suite uses algorithms designed for wrist worn accelerometers (Cole et al., 1992; Sadeh et al., 1994) and attempts to validate sleep time measurement algorithms using waist worn actigraphy in adults has been limited and demonstrated poor agreement with gold standard measurement (polysomnography) (Slater et al., 2015). Therefore, to compensate for participants who did not remove the accelerometer before sleeping nor kept a record of their sleep and wake times, data from 0h00 to 5h59 from each day were not analyzed. Wear time was calculated using Choi et al. (2011) algorithm. The accelerometer needed to register 600 min of wear time for a day to be considered a valid (Troiano et al., 2008). As the IPAQ-SF asks participants for time spent "sitting on a weekday," accelerometer data was averaged across all available valid weekdays. Three valid weekdays out of the 7-day period were required for participants' accelerometry data to be included in the analysis (Troiano et al., 2005, 2008).

Troiano et al. (2008) adult physical activity cut-off points were used to determine time spent engaging in SB ($\leq$99 counts/min) using the vector magnitude of all three axes. While the wGT3X + does have a built in inclinometer, it has not been shown to be better at detecting SB time than using count per minute thresholds in post-secondary students (Peterson et al., 2015). SB bouts were defined in 3 ways: $\geq$1 minute, $\geq$ 10 min, and $\geq$30 min. For all definitions, a tolerance of 2 min spent above the maximum counts per minute was permitted before a SB bout ended.

Agreement between measurements was assessed primarily through Bland-Altman (BA) mean-difference plots with 95% Limits of Agreement (LoA). Positive differences represented overestimation by the Week 4 IPAQ-SF assessment versus the Week 1 IPAQ-SF or accelerometry. Regression analysis was performed on the BA plots to determine if any bias between measurements may be fixed or proportional. Paired *t*-tests were used to determine if the mean bias between measurements represented a statistically significant difference. Based on ICC guidelines presented by Koo and Li (2016), retest reliability coefficients were calculated using an absolute agreement single measure, two-way mixed-effect model ICC ($ICC_{A,1}$). $ICC_{A,1}$ is reported for agreement between IPAQ and accelerometry as well. Furthermore, as Spearman's rank correlation coefficients is a commonly reported indicator of validity and reliability in the literature, it is also reported here for comparability to other studies (missing values excluded pairwise and confidence intervals calculated by applying Fisher's *z*-transformation) but is not a primary outcome of this study.

All statistical analyses were performed in RStudio (v1.1.453; R v3.4.4) with package 'psych' (v.1.8.4) installed (Revelle, 2017). Bonferroni corrections (α = 0.0167) were applied to account for the three definitions used to assess SB with the accelerometer.

3. Results

3.1. Demographics

Of 130 participants who enrolled in the parent study, 113 completed the study, and 103 participants had sufficient accelerometry data. At Week 1, 105 of 130 participants had valid IPAQ-SF sitting responses (did not select "Don't know/Not sure"). At Week 4, 84 of 113 participants had valid IPAQ-SF sitting responses, resulting in 74 participants with valid responses at both week 1 and 4. Meanwhile, 74 participants had both valid accelerometry data and IPAQ responses at week 4. Table 1 describes characteristics of the participant who had valid data, separated by whether data was available for the validity and reliability assessments as well as replicating overall sample data presented by Duncan et al. (2017). Fig. 1 illustrates the interrelationship of valid data collected.

Table 1
Summary of participant demographics separated by assessment.

Demographic	Enrolled (n = 130)	Completed (n = 113)	Validity analysis (n = 74)	Reliability analysis (n = 74)
Male:Female	80:50	68:45	47:27	48:26
Current smokers	63	55	33	33
Outpatient:Inpatient	124:6	108:5	70:4	70:4
Mean (*sd*) Age	40.1 (11.6) years	41.0 (11.7) years	42.3 (11.9) years	41.9 (11.8) years
Ethnicity				
African descent	21	19	12	12
Asian/South Asian	20	16	11	11
Caucasian and Hispanic	74	68	44	43
Multi-ethnic	5	5	3	3
Other	10	5	4	4
Symptom severity				
BPRS-A mean score (*sd*)	34.3 (7.6)	34.2 (8.5)	33.6 (7.1)	34.1 (9.0)
CGI-S mean score (*sd*)	3.5 (1.1)	3.5 (1.1)	3.4 (1.0)	3.4 (1.1)
AES mean score (*sd*)	32.0 (7.9)	31.7 (7.9)	30.6 (7.4)	31.8 (7.8)
CPZ equivalents mean (*sd*)	788.5 (1237.3) mg	748.1 (1216.6) mg	637.7 (819.9) mg	603.1 (820.4) mg
BMI*				
Mean (*sd*)	31.7 (8.1) kg/m^2	31.5 (8.4) kg/m^2	30.6 (7.2) kg/m^2	30.0 (6.1) kg/m^2
Underweight (BMI < 18.5)	1	1	1	1
Normal Weight (18.5 < BMI < 25)	21	19	13	13
Overweight (25 < BMI < 30)	35	32	24	24
Obese (BMI > 30)	72	60	35	35
Education				
Some high school (no diploma)	28	22	13	15
High school diploma	36	31	14	16
At least some Postsecondary	62	59	45	41
Other (e.g. apprenticeship)	4	2	2	2
Employment				
Full-time	3	2	2	2
Part-time	36	33	23	24
Student	7	6	3	4
Unemployed	77	65	38	35
Other (e.g. retired, volunteer)	7	7	8	8

Note: Values are frequency unless otherwise specified. Higher score represent more severe symptoms.

* One participant opted out of being weighed. BPRS-A = Brief Psychiatric Rating Scale 18-item Anchored version (Woerner et al., 1988), CGI-S = Clinical Global Impression Severity Scale (Guy, 1976), AES = Apathy Evaluation Scale (Marin et al., 1991), CPZ = Chlorpromazine Equivalents (Gardner et al., 2010), BMI = Body Mass Index.

3.2. IPAQ validity

Table 2 summarizes both the IPAQ-SF sitting scale (at Week 1 and 4) and average daily SB as measured by accelerometry for all valid responses, as well as Spearman's correlation between methods. $ICC_{A,1}$

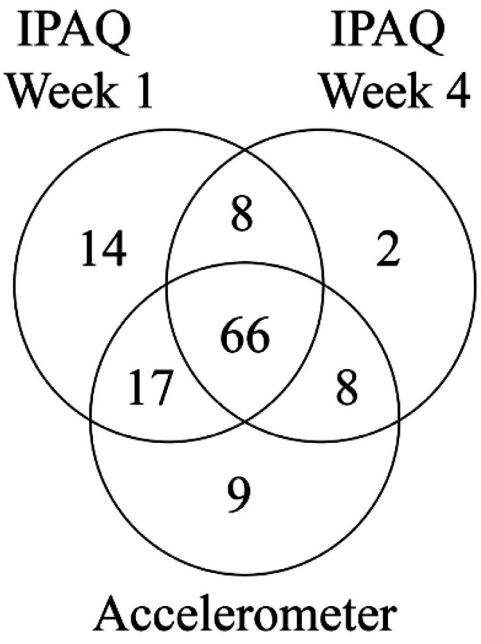

Fig. 1. Venn diagram representing overlap of sedentary behaviour data sources across participants. Participants with no physical activity data = 6.

between methods was lowest and non-significant for ≥30-min bouts: 0.09 (95% CI: −0.05–0.24), $p = 0.04$. $ICC_{A,1}$ between methods for ≥1-min and ≥10-min bouts were comparable at 0.21 (95% CI: 0.04–0.36), $p = 0.004$ and 0.23 (95% CI: 0.06–0.39), $p = 0.004$ respectively; and were statistically significant at the Bonferroni corrected alpha.

Fig. 2a–c includes BA plots of the difference between each accelerometry SB definition and Week 4 IPAQ-SF against the mean of both methods. Regression lines are displayed in Fig. 2a–c as alternating short and long dashed lines. Linear regression was significant and explained 32% of the variance when accelerometry SB was measured in ≥1-min bouts, 29% for ≥10-min bouts, and 45% for ≥30-min bouts. Mean difference between IPAQ-SF and accelerometry ≥ 1-min SB bouts was −86.1-min, 95% LoA: −569.9-min–397.6-min, Cohen's $d_z = 0.35$, $1 - \beta = 0.91$; and for ≥30-min SB bouts was 220.7 min, LoA: −266.6–708.0, Cohen's $d_z = 0.89$, $1 - \beta = 1.0$. T-tests revealed significant differences for both comparisons. Mean difference for ≥10-min bouts was 26.8 min, LoA: −458.7–512.3, Cohen's $d_z = 0.11$, $1 - \beta = 0.23$, but *t*-test results were not statistically significant. Results from statistical analyses for regressions and paired *t*-tests are summarized in Table 3.

3.3. IPAQ reliability

Fig. 2d includes the BA plot of the difference between Week 1 and Week 4 administrations of the IPAQ-SF against the mean of both time points. Week 1 and Week 4 administrations differed by an average of −26.6 min, 95% LoA: −564.2–510.9, Cohen's $d_z = 0.10$, $1 - \beta = 0.13$. but this was not statistically significant. Linear regression analysis of the plot was also not significant (results of statistical analyses also reported in Table 3), indicating that bias between measures is fixed. $ICC_{A,1}$ between time-points was 0.41 (95% CI: 0.21–0.59, $p < 0.001$), representing a 'fair' association between measurement times (Cicchetti, 1994).

4. Discussion

Previous evidence of validity suggests that the IPAQ-SF overestimates SB by 130 min per day with LoA of ± 6.75 h per day when using an accelerometry definition of 100 counts/min and no minimum bout length (Hagstromer et al., 2010). The comparable analysis performed in this study (bouts ≥ 1-min) suggests that the IPAQ-SF may underestimate by 86 min compared to accelerometry defined SB, but with even larger LoA (± 8 h). This is in line with Stubbs et al. (2016) hypothesis that self-report tends to underestimate SB among people with severe mental illness, despite the opposite being evident in the general population. One plausible explanation may be related to symptoms such as apathy and amotivation (e.g. less effortful recall; activities of daily life perceived as more effortful and thus not SB). However, more exploration as to why this is the case is warranted in order identify accurate self-report methods for collecting information on SB. However, using a minimum 10-min bout length seems to improve the mean bias so that it is no longer significantly different, but LoA remain large. As well, much like the data presented by Hagstromer et al. (2010) from the general adult population, all BA analyses suggested proportional bias occurs when using the IPAQ-SF.

With regards to correlational evidence, our results suggest that relationships between accelerometry and the IPAQ-SF in individuals with schizophrenia is on the low end of what is reported in the literature from the general adult population. ICCs between the accelerometer and the IPAQ-SF ranged from 0.09–0.23, and Spearman correlations ranged from 0.25–0.30, compared to Spearman correlations of 0.07–0.61 reported by Healy et al. (2011). The ICC as a measure of absolute agreement provides a better indication of whether the minute values between the two methods align, which given the agreement data presented in BA plots, is unsurprisingly low. Even Spearman's correlation – which is a better indicator of the general monotonic relationship (i.e. higher sitting scores on the scale are associated with higher levels of measured SB) – was low.

Table 2
Sedentary Behaviour over Two 1-Week periods.

Assessment	Descriptive statistics						Correlation with IPAQ Week 4	
	Mean	Median	*sd*	*n*	Min	Max	$ICC_{A,1}$	Spearman's rho
Validity								
IPAQ Week 4	424.6	390.0	247.0	74	0	960.0	–	–
Accelerometer:								
≥1-min Bouts	515.7	524.8	135.8	74	142.8	827.2	0.21 (CI: 0.04–0.36), $p = 0.004$**	ρ = 0.25, $p = 0.029$
≥10-min Bouts	403.1	408.6	138.1	74	99.0	748.8	0.23 (CI: 0.06–0.39), $p = 0.004$**	ρ = 0.30, $p = 0.009$**
≥30-min Bouts	213.6	190.4	111.9	74	25.0	490.8	0.09 (CI: −0.05–0.24), $p = 0.04$	ρ = 0.26, $p = 0.023$
Reliability								
IPAQ Week 1	447.4	360.0	260.1	74	60.0	1210.2	0.41 (CI: 0.21–0.59), $p < 0.001$*	ρ = 0.49, $p < 0.001$*
IPAQ Week 4	420.8	390.0	245.9	74	0.0	960.0	–	–

Note: IPAQ Week 4 descriptive statistics are reported separately for the individuals with valid accelerometry and with valid week 1 IPAQ responses. For correlations
* Statistically significant at α = 0.05 where appropriate.
** Statistically significant at Bonferroni corrected α = 0.0167 where appropriate.

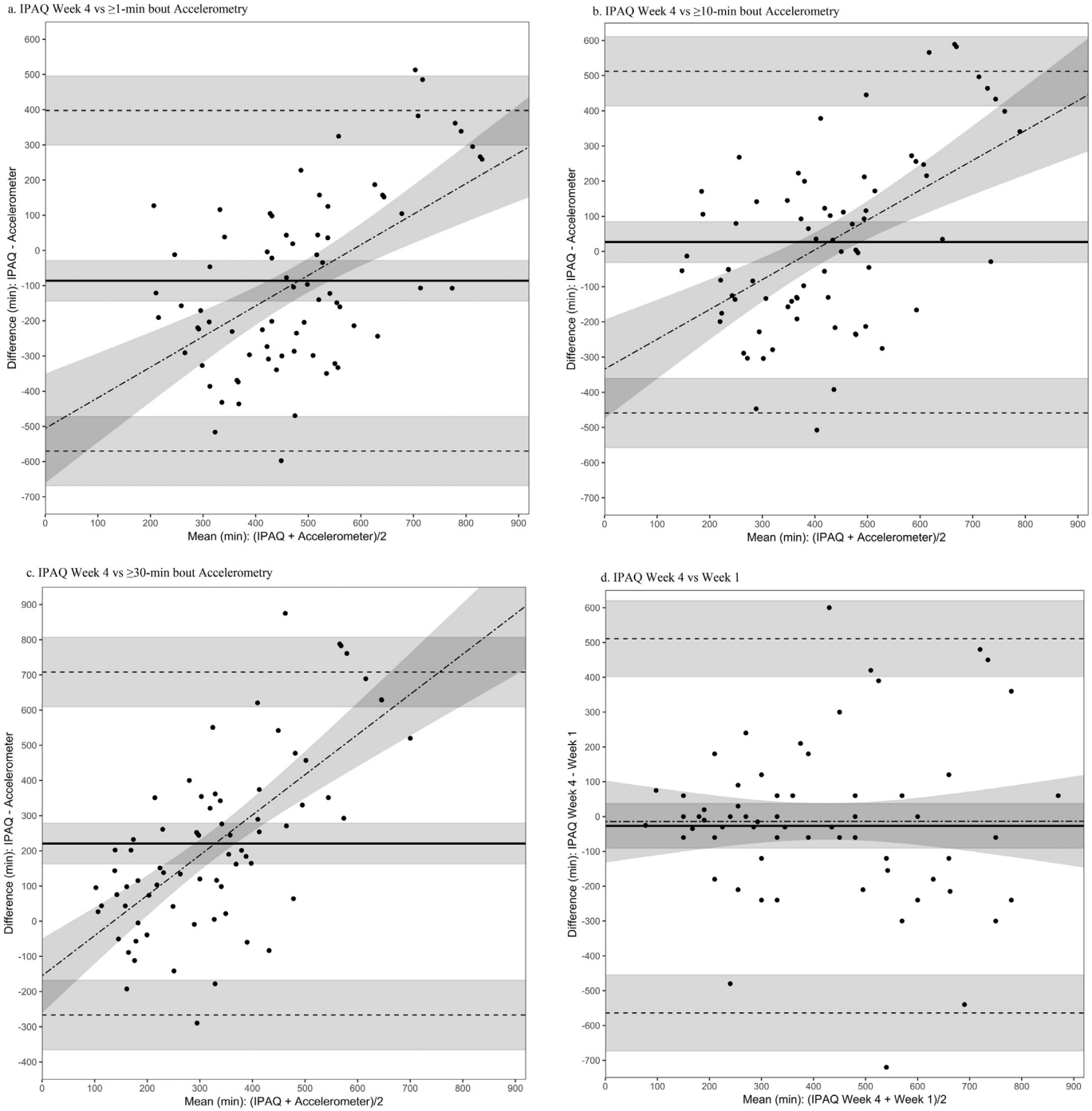

Fig. 2. a–d. Bland-Altman Plots of SED measured by IPAQ-SF and accelerometry. Solid lines indicate mean difference (bias), long dashed lines indicate 95% agreement limits, and alternating short and long dashes indicate the linear regression line. Grey areas represent 95% confidence intervals around respective lines.

Reliability indices were also quite low compared to the general adult population. The ICC reported here of 0.41 is well outside the range of 0.80–0.97 reported in the literature for the general adult population (Healy et al., 2011). Spearman's correlation did however fall within the range reported by Healy et al. (2011) ($\rho = 0.41$ vs $\rho = 0.18$–0.95) despite representing a 4-week span, though again, the ICC used here is a more appropriate indicator of whether the same values are reported at both time points.

Contrary to the hypothesis, the IPAQ-SF sitting scale best represented accelerometer derived SB when minimum bout lengths were set to ≥10-min, followed by ≥1-min, and finally ≥30-min bouts. While all three comparisons demonstrated statistically significant proportional bias such that higher amounts of SB resulted in more overestimation by the IPAQ-SF, this bias was most pronounced in the ≥30-min bout comparison where the slope of the regression line was 1.14 and explained 45% of the variance, compared to slopes of 0.85 and 0.86, and 29%–32% of the variance explained when using the smaller minimum bout lengths.

With regards to mean bias, IPAQ-SF scores overestimated by 28.7 min per day but did not differ significantly from SB bouts of ≥10-min, whereas the IPAQ-SF scores did significantly underestimate compared to SB bouts of ≥1-min by 84.0 min on average, and scores significantly overestimated when compared to SB bouts of ≥30-min by 224.7 min on average. Based on these findings, when using the IPAQ-SF

Table 3
Bland-Altman Plot Statistical Analyses.

Assessment	Regression analysis						Paired T-test			
	R^2	*df*	*F*	*p*	*β1*	*β0*	*Δ (sd)*	*t*	*df*	*p*
Validity										
≥1-min Bouts	0.32	1, 72	31.9	<0.001[a]	0.86	−505.5	−86.1 (246.8)	3.00	73	0.004[a]
≥10-min Bouts	0.29	1, 72	29.8	<0.001[a]	0.85	−334.2	26.8 (247.7)	.93	73	0.36
≥30-min Bouts	0.45	1, 72	59.9	<0.001[a]	1.14	−155.2	220.7 (248.6)	7.63	73	<0.001[a]
Reliability										
IPAQ Week 1	0.01	1, 72	0.28	0.6	−0.08	7.9	−26.6 (274.3)	0.84	73	0.41

[a] Significant with Bonferroni correction.

sitting scale as an indicator of SB among individuals with schizophrenia, researchers should interpret this as an indicator of the amount of SB engaged in bouts of at least 10 min – much like how the IPAQ-SF asks for bouts of PA of at least 10 min. Interpreting IPAQ-SF sitting data as an indicator of all SB time (i.e. ≥1-min bouts) may also be acceptable, but less accurate. However, the sitting scale should not be interpreted as an indicator of just prolonged sitting time (i.e. ≥30-min bouts).

It is unclear why the IPAQ performs best as an estimate of 10-minute bouts of sitting time, though several explanations are plausible. Foremost, the IPAQ PA questions ask for minimum 10-min bouts. While it does not do this for the sitting item, participants may be cued to assess their day for sitting in a similar fashion as when trying to recall PA. Additionally, it is possible that participants are able to factor in even short periods of SB to their estimate/recall such as a bus ride or waiting for an appointment. Research into the response processes used by participants to estimate PA and SB may help understand how activity profiling questions are responded to.

However, while the ≥10-min bout of SB interpretation does appear to be the best way to interpret IPAQ-SF sitting scale score, there are still substantial issues with the use of this self-report questionnaire as an indicator of SB. Firstly, the 95% LoA are quite large at approximately ± 8 h/day among all three bout definitions. Thus, while the bias did not differ significantly in the ≥10-min bout analysis on average, individual scores vary widely from accelerometry estimates. Secondly, the presence of proportional bias is a concern for validity as well, with overestimation tending to occur with higher levels of SB while underestimation occurs with lower levels. Finally, from a theoretical stand point, while the prelude to the sitting scale includes the instructions: "This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television," the question itself only asks about sitting. As such, SB in a lying or reclining position may not be considered by participants when responding to this question.

Repeated administration of the IPAQ-SF resulted in no significant mean bias between time-points, nor was there significant proportional bias. However, LoA were approximately ± 9 h/day, and the ICC was much lower than those reported in (Healy et al., 2011). Thus, while there appears to be no bias between the two measurement periods, there was substantial inconsistency. However, this inconsistency may represent actual differences in individual sitting behaviours between the two measurement periods. Future studies aiming to verify the reliability of physical activity and SB measurement tools should consider methods to address this possibility, such as using objective measures to assess the behaviours over both periods to determine if the behaviour was actually similar (Duncan et al., 2017), or by shortening the retest duration so that both administration cover similar time periods (e.g. two administrations on the same day, or one day apart so that the majority of days in the 7-day measurement period are covered).

The results presented here suggest that the IPAQ-SF sitting scale may be an unsuitable measure of SB for many scenarios when working with individuals with schizophrenia including at a population level or with smaller samples. While mean values do compare to objective accelerometry, other psychometric properties are lacking. LoA are large and bias is proportional, instilling a lack of confidence in whether the actual "minutes" of activity reported by the IPAQ-SF reflect the minutes of SB engaged in by any one *individual*. As a result, any correlational data based on the IPAQ-SF or cut points based on the number of minutes reported should be interpreted with caution. Given low Spearman correlations, even interpretations about who has high vs low rates of SB within a sample cannot be made with confidence. At best, the IPAQ-SF can be used to indicate the mean SB of a *sample.* As Firth and colleagues (Firth et al., 2018) suggest regarding the PA scales, similar caution is likely required if comparing sitting time measured by the IPAQ-SF among individuals with schizophrenia and control. Given that BA analysis suggests similar characteristics (i.e., mean and proportional biases) when used in the general adult population (Hagstromer et al., 2010) the IPAQ-SF may be adequate at best when comparing to the general adult population. However, any further data generated should be considered preliminary or exploratory and would require replication with more objective methods in order to confirm whether any relationship exists between SB and a proposed outcome or predictor.

To our knowledge this is the first study to generate evidence of validity when using the IPAQ-SF sitting scale in a sample of individuals with schizophrenia. However, as a secondary analysis, some limitations were inherent. The use of accelerometers without the additional use of a purpose-built inclinometer only allows the use of counts per minute to classify SB but does not take into account posture. Thus, behaviours such as standing with low activity counts per minute may be improperly classified as SB.

Unfortunately, much data were unusable in this sample. While accelerometer adherence was high (only 8.9% of participants had insufficient wear time), IPAQ responses that could be compared against the accelerometry were minimal. At Week 4, 25.7% of the completing sample (and 29.5% of those with valid accelerometry) selected "Don't know/Not sure" option on the sitting item, which could not be compared to accelerometry as a result. Thus, the usable sample was reduced to 74. Given the large effect size observed for SB bouts ≥1 and ≥30 min, these analyses were reasonably powered ($1-\beta \geq 0.91$ when $\alpha = 0.05$) to avoid both Type I and Type II errors. However, with the small effect observed for SB bouts ≥10-min a false negative may have resulted (i.e. no statistical difference between measures concluded despite a difference existing). That being said, whether the IPAQ differs *statistically* from accelerometry, may be less relevant than simply characterizing the relationship between the resultant measurements to indicate the implications of using the IPAQ as an assessment of SB. Furthermore, while this is a small sample compared to the initial intention of the IPAQ (Bauman et al., 2009a, 2009b; Craig et al., 2003), results are of a similar magnitude as reported by Hagstromer et al. (2010) from larger sample of the general population.

Additionally, while this study builds on previous efforts to develop validity evidence of the IPAQ-SF when used with samples of individuals with schizophrenia (Duncan et al., 2017; Faulkner et al., 2006), the

focus of these efforts has been predominantly on the evidence of convergence (American Educational Research Association et al., 2014) with other measurement methods, that is the accuracy of the tool relative to other methods. However, there are several other sources of evidence for validity that can also be considered, namely evidence based on: test content, response processes, internal structure, and the consequences of testing (AERA, APA, NCME, & JCSEPT, 2014). Having analyzed the accelerometry data in varying minimum bout lengths provides some insight on response processes (i.e., how participants approach responding to items (Padilla and Benítez, 2014; AERA, APA, NCME, & JCSEPT, 2014)). Even short bouts of sitting appear to be considered when individuals with schizophrenia respond to the IPAQ-SF. However, other methods of assessing response processes such as eye-tracking (to identify what participants are attending to when presented with a questionnaire) and cognitive interviews (where participants are asked to describe aloud how they are interpreting the items, and their process behind generating a response) (Padilla and Benítez, 2014) may help inform the interpretation of IPAQ-SF data and help researchers design more accurate self-report questionnaires for intervention and surveillance work. In particular, identifying why participants opt to select "Don't know/Not sure" may help reduce lost data.

Funding

This study was supported by a Canadian Institutes of Health Research (CIHR) operating grant #MOP-115709. Guy Faulkner is supported by a Canadian Institutes of Health Research-Public Health Agency of Canada Chair in Applied Public Health.

Contributors

Dr. Faulkner, Dr. Arbour-Nicitopoulos, and Dr. Remington designed the study. Markus Duncan and Mehala Subramaniapillai performed data collection. Markus Duncan performed data analysis and prepared the manuscript.

Conflicts of interest

Dr. Remington has received consultant fees from Neurocrine Biosciences, Novartis, and Synchroneuron, as well as grant support from Novartis. All other authors declare that they have no conflicts of interest.

Acknowledgements

The authors would like to thank Carol Borlido for her role in recruiting participants to the study and collecting data.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing, 2014. Standards for Educational and Psychological Testing. American Educational Research Association.

Appelbaum, P.S., Grisso, T., 2001. MacCAT-CR: MacArthur Competence Assessment Tool For Clinical Research. Professional Resource Press, Sarasota, FL.

Arbour-Nicitopoulos, K.P., Duncan, M.J., Remington, G., Cairney, J., Faulkner, G.E., 2017. The utility of the health action process approach model for predicting physical activity intentions and behavior in schizophrenia. Front. Psychiatry 8, 1–8. https://doi.org/10.3389/fpsyt.2017.00135.

Bauman, A., Ainsworth, B.E., Bull, F., Craig, C.L., Hagströmer, M., Sallis, J.F., Pratt, M., Sjöström, M., 2009a. Progress and pitfalls in the use of the International Physical Activity Questionnaire (IPAQ) for adult physical activity surveillance. J. Phys. Act. Health 6, S5–S8. https://doi.org/10.1123/jpah.6.s1.s5.

Bauman, A., Bull, F., Chey, T., Craig, C.L., Ainsworth, B.E., Sallis, J.F., Bowles, H.R., Hagstromer, M., Sjostrom, M., Pratt, M., Díaz, C.G., Bazan, N., Kunic, H., Bauman, A., Merom, D., Smith, B., De Bourdeaudhuij, I., Lefevre, J., Philippaerts, R., Matsudo, S.M., Matsudo, V.R., Craig, C.L., Cameron, C., Yang, L., Hua, F., Gómez, L.F., Fromel, K., Mitas, J., Macfarlane, D., Bacon-Shone, J., Reddy, S.K., Joshi, P., Goenka, S., Prabhakaran, D., Katsumura, T., Murase, N., Volbekiene, V., Baubliene, R., McLean, G., Carr, H., Tomten, H., Anderssen, S.A., Sardinha, L., Mota, J., Al-Hazzaa, H.M., Serra Majem, L., Roman, B., Sjöström, M., Hagströmer, M., Bergman, P., Yiing, M.L., Yung-Tai, H., Ainsworth, B.E., Hipp, D., 2009b. The international prevalence study on physical activity: results from 20 countries. Int. J. Behav. Nutr. Phys. Act. 6, 1–11. https://doi.org/10.1186/1479-5868-6-21.

Biswas, A., Oh, P.I., Faulkner, G.E., Bajaj, R.R., Silver, M.A., Mitchell, M.S., Alter, D.A., 2015. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults a systematic review and meta-analysis. Ann. Intern. Med. 162, 123–132. https://doi.org/10.7326/M14-1651.

Bueno-Antequera, J., Oviedo-Caro, M.Á., Munguía-Izquierdo, D., 2017. Relationship between objectively measured sedentary behavior and health outcomes in schizophrenia patients: the PsychiActive project. Schizophr. Res. https://doi.org/10.1016/j.schres.2017.11.022.

Canadian Society of Exercise Physiology, 2012. Canadian Physical Activity and Sedentary Behaviour Guidelines 1–29. doi: 978-1-896900-30-8.

Choi, L., Liu, Z., Matthews, C.E., Buchowski, M.S., 2011. Validation of accelerometer wear and nonwear time classification algorithm. Med. Sci. Sports Exerc. 43, 357–364. https://doi.org/10.1249/MSS.0b013e3181ed61a3.

Cicchetti, D.V, 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290. https://doi.org/10.1037/1040-3590.6.4.284.

Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J., Gillin, J.C., 1992. Automatic sleep/wake identification from wrist activity. Sleep 15, 461–469.

Correll, C.U., Solmi, M., Veronese, N., Bortolato, B., Rosson, S., Santonastaso, P., Thapa-Chhetri, N., Fornaro, M., Gallicchio, D., Collantoni, E., Pigato, G., Favaro, A., Monaco, F., Kohler, C., Vancampfort, D., Ward, P.B., Gaughran, F., Carvalho, A.F., Stubbs, B., 2017. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 16, 163–180. https://doi.org/10.1002/wps.20420.

Craig, C.L., Marshall, A.L., Sjöström, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F., Oja, P., 2003. International physical activity questionnaire: 12-country reliability and validity. Med. Sci. Sports Exerc. 35, 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB.

Dixon, L., Weiden, P., Delahanty, J., Goldberg, R., Postrado, L., Lucksted, a, Lehman, a, 2000. Prevalence and correlates of diabetes in national schizophrenia samples. Schizophr. Bull. 26, 903–912.

Duncan, M.J., Arbour-Nicitopoulos, K., Subramanieapillai, M., Remington, G., Faulkner, G., 2017. Revisiting the International Physical Activity Questionnaire (IPAQ): assessing physical activity among individuals with schizophrenia. Schizophr. Res. 179, 2–7. https://doi.org/10.1016/j.schres.2016.09.010.

Faulkner, G.E., Cohn, T., Remington, G., 2006. Validation of a physical activity assessment tool for individuals with schizophrenia. Schizophr. Res. 82, 225–231.

Firth, J., Stubbs, B., Vancampfort, D., Schuch, F.B., Rosenbaum, S., Ward, P.B., Firth, J.A., Sarris, J., Yung, A.R., 2018. The validity and value of self-reported physical activity and accelerometry in people with schizophrenia: a population-scale study of the UK Biobank. Schizophr. Bull. 44, 1293–1300. https://doi.org/10.1093/schbul/sbx149.

Gardner, D.M., Pharm, D., Murphy, A.L., Pharm, D., Donnell, H.O., Pharm, B.S., Centorrino, F., Baldessarini, R.J., Donnell, O., Al, E.T., 2010. International Consensus Study of Antipsychotic Dosing 686–693. https://doi.org/10.1176/appi.ajp.2009.09060802.

Guy, W., 1976. Clinincal global impressions. ECDEU Assessment Manual for Psychopharmacology-Revised National Institute of Mental Health, Rockville, MD, pp. 218–222.

Hagstromer, M., Ainsworth, B.E., Oja, P., Sjostrom, M., 2010. Comparison of a subjective and an objective measure of physical activity in a population sample. J. Phys. Act. Health 7, 541–550.

Healy, G.N., Clark, B.K., Winkler, E.A.H., Gardiner, P.A., Brown, W.J., Matthews, C.E., 2011. Measurement of adults' sedentary time in population-based studies. Am. J. Prev. Med. 41, 216–227. https://doi.org/10.1016/j.amepre.2011.05.005.

Hennekens, C.H., Hennekens, A.R., Hollar, D., Casey, D.E., 2005. Schizophrenia and increased risks of cardiovascular disease. Am. Heart J. https://doi.org/10.1016/j.ahj.2005.02.007.

Keefe, R.S.E., Fenton, W.S., 2007. How should DSM-V criteria for schizophrenia include cognitive impairment? Schizophr. Bull. 33, 912–920. https://doi.org/10.1093/schbul/sbm046.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Manu, P., Dima, L., Shulman, M., Vancampfort, D., De Hert, M., Correll, C.U., 2015. Weight gain and obesity in schizophrenia: Epidemiology, pathobiology, and management. Acta Psychiatr. Scand. 97–108. https://doi.org/10.1111/acps.12445.

Marin, R.S., Biedrzycki, R.C., Firinciogullari, S., 1991. Reliability and validity of the Apathy Evaluation Scale. Psychiatry Res. 38, 143–162.

McNamee, L., Mead, G., MacGillivray, S., Lawrie, S.M., 2013. Schizophrenia, poor physical health and physical activity: evidence-based interventions are required to reduce major health inequalities. Br. J. Psychiatry. https://doi.org/10.1192/bjp.bp.112.125070.

Owen, N., Healy, G.N., Matthews, C.E., Dunstan, D.W., 2010. Too much sitting: the population health science of sedentary behavior. Exerc. Sport Sci. Rev. 38, 105–113. https://doi.org/10.1097/JES.0b013e3181e373a2.

Padilla, J.-L., Benítez, I., 2014. Validity evidence based on response processes. Psicothema 26, 136–144. https://doi.org/10.7334/psicothema2013.259.

Peterson, N.E., Sirard, J.R., Kulbok, P.A., Deboer, M.D., Erickson, J.M., 2015. Validation of accelerometer thresholds and inclinometry for measurement of sedentary behavior

in young adult university students. Res. Nurs. Health 38, 492–499. https://doi.org/10.1002/nur.21694.
Reichenberg, A., Weiser, M., Caspi, A., Knobler, H.Y., Lubin, G., Harvey, P.D., Rabinowitz, J., Davidson, M., 2006. Premorbid intellectual functioning and risk of schizophrenia and spectrum disorders. J. Clin. Exp. Neuropsychol. 28, 193–207. https://doi.org/10.1080/13803390500360372.
Revelle, W.R., 2017. psych: Procedures for personality and psychological research.
Rosenbaum, S., Ward, P.B., 2016. The Simple Physical Activity Questionnaire. Lancet Psychiatry 3, e1. https://doi.org/10.1016/S2215-0366(15)00496-4.
Sadeh, A., Sharkey, K.M., Carskadon, M.A., 1994. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 17, 201–207.
Sallis, J.F., Owen, N., Fotheringham, M.J., 2000. Behavioral epidemiology: a systematic framework to classify phases of research on health promotion and disease prevention. Ann. Behav. Med. 22, 294–298. https://doi.org/10.1007/BF02895665.
Sheehan, D.V, Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 2), 22–57.
Slater, J.A., Botsis, T., Walsh, J., King, S., Straker, L.M., Eastwood, P.R., 2015. Assessing sleep using hip and wrist actigraphy. Sleep Biol. Rhythms 13, 172–180. https://doi.org/10.1111/sbr.12103.
Stubbs, B., Williams, J., Gaughran, F., Craig, T., 2016. How sedentary are people with psychosis? A systematic review and meta-analysis. Schizophr. Res. 171, 103–109. https://doi.org/10.1016/j.schres.2016.01.034.
Thorp, A.A., Owen, N., Neuhaus, M., Dunstan, D.W., 2011. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996–2011. Am. J. Prev. Med. 41, 207–215. https://doi.org/10.1016/j.amepre.2011.05.004.
Tremblay, M.S., Aubert, S., Barnes, J.D., Saunders, T.J., Carson, V., Latimer-Cheung, A.E., Chastin, S.F.M., Altenburg, T.M., Chinapaw, M.J.M., 2017. Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. Int. J. Behav. Nutr. Phys. Act. 14, 75. https://doi.org/10.1186/s12966-017-0525-8.
Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T., McDowell, M., 2008. Physical activity in the United States measured by accelerometer. Med. Sci. Sport Exerc. 40, 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3.
Trost, S.G., McIver, K.L., Pate, R.R., 2005. Conducting accelerometer-based activity assessments in field-based research. Med. Sci. Sport Exerc. 37, S531–S543. https://doi.org/10.1249/01.mss.0000185657.86065.98.
Vancampfort, D., Correll, C.U., Galling, B., Probst, M., De Hert, M., Ward, P.B., Rosenbaum, S., Gaughran, F., Lally, J., Stubbs, B., 2016a. Diabetes mellitus in people with schizophrenia, bipolar disorder and major depressive disorder: a systematic review and large scale meta-analysis. World Psychiatry 15, 166–174. https://doi.org/10.1002/wps.20309.
Vancampfort, D., De Hert, M., De Herdt, A., Soundy, A., Stubbs, B., Bernard, P., Probst, M., 2014. Associations between perceived neighbourhood environmental attributes and self-reported sitting time in patients with schizophrenia: a pilot study. Psychiatry Res.
Vancampfort, D., Firth, J., Schuch, F.B., Rosenbaum, S., Mugisha, J., Hallgren, M., Probst, M., Ward, P.B., Gaughran, F., De Hert, M., Carvalho, A.F., Stubbs, B., 2017. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. World Psychiatry 16, 308–315. https://doi.org/10.1002/wps.20458.
Vancampfort, D., Probst, M., Knapen, J., Carraro, A., De Hert, M., 2012. Associations between sedentary behaviour and metabolic parameters in patients with schizophrenia. Psychiatry Res. 200, 73–78.
Vancampfort, D., Rosenbaum, S., Probst, M., Connaughton, J., du Plessis, C., Yamamoto, T., Stubbs, B., 2016b. What are the top 10 physical activity research questions in schizophrenia? Disabil. Rehabil. 8288, 1–9. https://doi.org/10.3109/09638288.2015.1116622.
Vancampfort, D., Stubbs, B., Mitchell, A.J., De Hert, M., Wampers, M., Ward, P.B., Rosenbaum, S., Correll, C.U., 2015. Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. World Psychiatry 14, 339–347. https://doi.org/10.1002/wps.20252.
Warren, T.Y., Barry, V., Hooker, S.P., Sui, X., Church, T.S., Blair, S.N., 2010. Sedentary behaviors increase risk of cardiovascular disease mortality in men. Med. Sci. Sports Exerc. 42, 879–885. https://doi.org/10.1249/MSS.0b013e3181c3aa7e.
Woerner, M.G., Mannuzza, S., Kane, J.M., 1998. Anchoring the BPRS: an aid to improved reliability. Psychopharmacol. Bull. 24, 112–117.