



# Can consumer grade activity devices replace research grade actiwatches in youth mental health settings?

Jan Scott<sup>1,2,3,4</sup> · Ashlee Grierson<sup>5</sup> · Lillian Gehue<sup>2</sup> · Havard Kallestad<sup>3,4</sup> · Iain MacMillan<sup>6</sup> · Ian Hickie<sup>2</sup>

Received: 9 October 2018 / Accepted: 24 December 2018  
© The Author(s) 2019

## Abstract

Mental health services increasingly recognize the benefits of monitoring sleep and activity patterns in youth. Objective, ecological monitoring via research grade actiwatches is valid, but an emerging literature suggests using cheaper, consumer grade activity tracking devices. This study aimed to compare the cross-validity of these devices, and their feasibility and usability in youth mental health settings. A convenience sample of 13 individuals recruited from youth mental health settings wore a consumer grade and research grade activity device for seven consecutive days. The five most frequently reported sleep and activity variables extracted directly from each device were compared using paired *t* tests, mean average percentage errors, and magnitude and direction of biases. Ten individuals provided usable recordings of sleep and activity parameters. Compared to the actiwatch recordings (criterion values), the consumer grade device showed systematic biases in the recording of total sleep time (overestimated by about an hour) and waking after sleep onset (underestimated by about half an hour). The consumer grade device was most reliable for monitoring daytime sedentary behaviour. However, participant feedback indicated that youth preferred the consumer grade device. The findings suggest that it is inappropriate to substitute actiwatches with commercial activity trackers in research setting, as the latter are probably insufficient for precision diagnostics or stratification of cases into treatment relevant subgroups. Consumer grade devices are more acceptable to youth, so they may be clinically useful for monitoring intra-individual where absolute accuracy is not required.

**Keywords** Youth · Actigraphy · Consumer tracker · Research device · Bland–Altman plots

## Abbreviations

BMI	Body mass index
ConD	Consumer grade sleep-activity tracking device
IQ	Intelligence quotient
IQR	Interquartile range
kg/m <sup>2</sup>	Kilograms per square metre

MAPE	Mean average percentage error
MVPA	Moderate or vigorous physical activity
ResD	Research grade actigraphy device
SE	Sleep efficiency
s.d.	Standard deviation
s.d.-diff	Standard deviation of the difference
SPSS	Statistical Package for the Social Sciences
TST	Total sleep time
WASO	Waking after sleep onset
YMH	Youth mental health

✉ Jan Scott  
jan.scott@newcastle.ac.uk

<sup>1</sup> Academic Psychiatry, Institute of Neuroscience, Newcastle University, Newcastle, UK

<sup>2</sup> Brain and Mind Centre, The University of Sydney, Sydney, Australia

<sup>3</sup> Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

<sup>4</sup> Department of Research and Development, St. Olavs University Hospital, Trondheim, Norway

<sup>5</sup> Clinical Research Unit for Anxiety and Depression, St Vincents Hospital, Sydney, Australia

<sup>6</sup> Peninsula Health, Frankston, VIC, Australia

## Introduction

Actigraphy is established as a valuable tool for research into sleep, physical and mental disorders. Actigraphy provides an ecologically valid assessment of sleep with a modest, but manageable loss of data reliability compared to polysomnography [1], and it is an acceptable measure of energy expenditure if indirect calorimetry is unfeasible [2]. Objective monitoring of both sleep and daytime activities is

preferable to self-report as the latter show significantly lower levels of reliability. For example, whilst adults self-reported that they spent 38% of their time in moderate or vigorous physical activity (MVPA), actigraphic recordings revealed that they spent only 5% of their time at this intensity [3]. Overall, there is growing recognition that actigraphy can be applied to real-time monitoring of free-living activities in general population and clinical studies [4].

Actigraphy is increasingly employed in research in mental disorders, including in individuals at risk of developing mood or bipolar disorders [5–7]. Evidence demonstrates that actiwatches have utility in evaluating sleep-wake cycles in clinical or natural environments, can be employed prospectively to monitor longitudinal changes in symptoms or the evolution of an illness prodrome and might be employed to assess response to clinical interventions or patterns of recovery [8, 9]. This has led to suggestions that actiwatches could be used more widely in day-to-day practice, especially in adolescents and young adult populations where both irregular sleep patterns and low daytime activity may trigger or exacerbate health problems [4, 10]. However, research grade actiwatches, with their required device readers and software programmes, are relatively expensive. This, plus the fact the actiwatches may be lost or not returned to the clinic, may prohibit their routine use in youth mental health (YMH) settings. Unsurprisingly, it has been suggested that the rapid evolution of technology may offer opportunities to employ cheaper, consumer-based monitors in research and clinical practice [11].

The use of commercially available fitness tracking devices and other wearable technology is a key emerging trend of the last decade [11]. This ‘quantified self’ movement has been made possible by the development of consumer acceptable, low-cost, stream-lined alternatives to research grade medical devices such as actigraphic watches [12–14]. As such, it is now considered the norm for an individual to track their daily activity levels and sleep patterns via an on-screen display on a small wearable or transportable gadget. These devices usually link to a web-based interface that allows summary data (e.g. weekly number of steps) to be displayed on a smart phone, tablet or computer [13]. However, it is unclear whether commercial grade tracking devices could or should be used as an alternative to research grade actiwatches in YMH settings. Critically, using a commercial device to record rest-activity patterns and monitor personal wellness or lifestyle behaviours is a different proposition from monitoring putative clinical markers of mental disorders. This is especially likely to be true if the latter results in the individual receiving a clinical diagnosis or a recommendation for treatment. In addition, despite widespread adoption of commercial devices by youth populations, there is relatively little information on their accuracy in clinical settings and there has been limited testing of the validity of

the various rest-activity outputs against ‘criterion’ measures [11, 15, 16].

This study examines the validity, feasibility and acceptability of using a consumer grade activity device (ConD) as a substitute for a commonly used research grade actigraphic device (ResD) in the measurement of the five sleep and activity variables that are reported most frequently in actigraphy studies in youth (see, e.g. [1, 6]). The specific aims were to:

- (a) compare seven consecutive days of 24-h recordings of sleep-wake cycles and to assess the cross-validity of the ConD with the ResD on five metrics associated with mental and physical health, namely sleep duration, waking after sleep onset, sleep efficiency, minutes spent in MVPA and proportion of time sedentary.
- (b) determine the device feasibility and usability by assessing study dropout and reviewing participant feedback regarding the use of ConD and ResD.

## Methods

Ethical approval was granted by the north east branch of the National Research and Ethical Committee (NREC) in the UK (reference: 12/NE/0325) and the University of Sydney Human Research Ethics Committee (HREC: 2015/4961) in Australia.

Prior to commencing recruitment to the comparison study, a literature review was undertaken to identify the time duration required for recordings [17] and the most appropriate wrist-worn ConD, including acceptable battery life and inter-device reliability of the chosen ConD [18–23]. (further information regarding the device selection process is available from the corresponding author). This established that Fitbit© devices were the most appropriate for the present study with good inter-device reliability between different models and ease of access to minute-level data if required (via a Fitbit application programming interface; Fitabase). The reference device (Actiwatch-64; Philips Respironics, USA) was selected as it was the actiwatch that had been used in youth studies undertaken by these and other researchers.

Adherence to the study protocol, recruitment of participants and collection of data was overseen at both sites by a senior researcher (JS).

## Sample

A convenience sample was recruited via clinicians involved in clinical and research programmes that undertook actigraphic monitoring of sleep and/or activity in youth (e.g. a pilot study for a new therapy for individuals at risk of bipolar disorders; a study of subjective and objective measures of

sleep and mood problems in YMH clinic attendees). Eligibility criteria were as used in previous studies by several previous research studies by our research groups (see, e.g. [6, 10]).

The inclusion criteria were that the individual was (a) aged 16–25 years and (b) willing and able to give written informed consent to participate. The exclusion criteria were: (1) clinically assessed IQ < 70, evidence of intellectual impairment and/or history of head injury; (2) mental disorder secondary to a medical condition; (3) substance or alcohol use disorder; (4) elevated risk of suicide or self-harm; (5) regular use of medications that affect sleep, melatonin secretion, circadian rhythms or alertness; (6) evidence of other sleep (e.g. sleep apnoea, narcolepsy), neurological (e.g. epilepsy) or primary medical conditions associated with sleep-wake dysfunction; (7) recent trans-meridian travel (i.e. potential for jet lag) or regular shift work; and (8) the presence of mobility problems (i.e. unable to walk unaided, etc.).

### Sleep and activity metrics

We identified the five most commonly reported markers of physical and mental health that were available from the ResD and could be extracted from the ConD without any additional input from the study participant or any need for further calculations (e.g. this excluded sleep onset latency as this requires additional information). There were three sleep and two activity metrics:

- (a) total sleep time (TST) and waking after sleep onset (WASO) in minutes; and sleep efficiency (SE; reported on a 0–1 value) as extracted from the device recordings.
- (b) time spent in physical activities of different intensities is reported directly by the ConD and by applying published algorithms for the ResD [24, 25]. The selected metrics, which are easily interpretable and have established positive or negative associations with health [3, 26], were: MVPA (time spent in active minutes) and Sedentary Behaviour (reported as a 0–1 value which represents the number of minutes sedentary divided by the monitoring time).

### Procedure

A researcher gave discussed the study protocol and completed the consent procedures with the participant, then basic demographic and health information was recorded or estimated (e.g. height and weight were used to calculate body mass index: BMI). Each participant received instructions on how to use the ResD and ConD (as required). Individuals who did not own a Fitbit© were provided with the basic model and a charger for the duration of the study. Time was

synchronized on both devices and individuals were asked to wear the ConD and ResD concurrently for seven consecutive days and nights on their non-dominant wrist (For the ConD, participants were reminded to recharge the unit on day 5). The ‘normal’ setting was selected for detection of sleep and activity on the ConD. For the ResD, the ‘usual’ threshold (medium sensitivity) was selected for sleep-wake detection.

At the end of the monitoring, the researchers collected data from the ConD for each study parameter using the same epochs as the ResD (ConD data were obtained by the participant downloading reports from the website and giving the data to a researcher; by giving the researchers direct access to the ConD or raw data from Fitabase; or by returning the ConD to a researcher). The participant returned the ResD to a researcher who downloaded and extracted recordings for each study parameter from the actiwatch. The two sets of recordings were combined into a data file and any personal identifying information was removed.

Individuals who commenced the monitoring week were asked to provide verbal feedback about their views and/or preferences for using a ConD or ResD. The four questions covered acceptability of wearing each device (in terms of, e.g. being seen wearing by their peers); how intrusive it was to wear and manage the device; any preference in terms of using or interacting with the device; any other personal comments or feedback.

### Statistical analyses

All analyses were planned a priori and undertaken using SPSS (version 23). In free-living conditions, it is known that adolescent sleep-wake patterns with entrainment (weekdays with regular scheduled activities) may differ from those with reduced or no entrainment (e.g. weekends), so analyses take this into account (see below).

Normality of distributions for all measured variables was established using the Shapiro–Wilk test and statistical significance was set at  $p < 0.05$ .

Descriptive statistics were used to characterize the sample; the ConD and ResD data were then compared using three approaches:

1. Paired samples *t* tests were used to assess systematic differences between the recordings obtained from the ConD and ResD for five consecutive weekdays/week-nights.
2. Mean absolute percentage error (MAPE) values were then calculated. The MAPE provides an indication of the absolute value of the error and is estimated for the ConD by dividing the absolute bias (ResD–ConD) by the ResD (criterion value) measure and multiplying by 100 [11, 13].

3. Bland–Altman (difference) plots were used to determine whether the ConD provides an over- or under-estimation of any metric compared to the ResD [27]. We plotted two points on each Bland–Altman graph for each participant (one represents the comparison of weekday values and the other the weekend values). To create the graph, it is first necessary to compute the mean bias for the measure (i.e. the mean difference between the ResD and ConD), along with the standard deviation (s.d.) of the bias. Next the lower and upper limits of the level of agreement are calculated ( $\pm 1.96 \times$  s.d. of the bias). The mean difference is then plotted against the average of the two measures (i.e. ConD plus ResD divided by two). A positive bias indicates that the ConD over-estimates the ResD values, whilst a negative bias indicates that the ConD under-estimates the ResD values.

### Power calculation

We used data from our own and other previous studies to determine the sample size required for the paired *t* tests (using <http://samplesizecalculator.com>). Assuming a TST of 7–8 h and a 10% difference between TST recorded on the ConD and ResD (42–48 min) and a predicted s.d. of the difference (s.d.-diff) of about 30 min, then 7–8 individuals are required to achieve 80% statistical power for identifying a statistically significant difference (at  $p < 0.05$ ). Likewise, six participants are required if we assume the SE is 0.8–0.9, with a 10% difference in the SE between devices (and a s.d.-diff of 0.05). Based on previously reported dropout rates from actigraphy (10–35%) and objective monitoring studies of youth (20–40%) and to allow for random missing recordings of sleep or activity variables, we estimated that if a minimum of 12 individuals commenced the study, we would obtain the required data from 8 to 10 individuals.

## Results

Thirteen individuals commenced data collection, but two did not attend the follow-up appointment (and did not provide data) and ConD data were unavailable for another individual.

The included sample comprised of 10 youth (6 females) with a median age of 19.3 years (interquartile range, IQR 17–21) and a median BMI of 22.3 kg/m<sup>2</sup> (IQR 18.9–25.7). All individuals reported depressive symptoms, four also reported anxiety symptoms, three reported symptoms of hypomania and one had a history of hallucinations. Five individuals had a family history of mood disorders (unipolar and/or bipolar) and four individuals were currently prescribed psychotropic medication. Characteristics of individuals recruited in England ( $n = 6$ ) did not differ from those recruited in Australia.

All ten individuals provided sleep and activity data for five consecutive weekdays, but one individual did not wear an actiwatch for the weekend. As shown in the paired *t* tests reported in Table 1, the ConD gives significantly higher values for weeknight TST and SE and for weekday MVPA compared to the ResD and significantly lower values for weeknight WASO and weekday sedentary behaviour. The MAPE indicates that the percentage error for measurements undertaken by the ConD is particularly high for WASO (45%), exceeds 10% for TST and MVPA, but is lower for the SE (9%) and Sedentary Behaviour (5%).

As shown in Table 2 and Fig. 1, the above findings translate into a high level of systematic bias for sleep measures, with the ConD over-estimating the TST by about one hour (+56.41 min) and underestimating WASO by about half an hour (–29.65 min) compared to the ResD. In addition, the ConD overestimated SE by about seven percent (+0.074). Daytime activity parameters were less prone to bias (MVPA overestimated by about 6 min; Sedentary Behaviour underestimated by about 4%) (Fig. 5).

**Table 1** Paired *t* tests for five selected health markers measured on consecutive weekdays using a commercial grade (ConD) and research grade device (ResD) and the estimated mean average percentage error of the ConD

Variable	Mean (s.d.)		Paired <i>t</i> test		MAPE (%)
	Consumer device	Research device	<i>t</i> test ( <i>df</i> 9)	<i>p</i>	
TST in minutes	482.21 (83.91)	425.30 (62.27)	13.52	0.001	13
WASO in minutes	35.98 (20.14)	65.63 (31.49)	5.84	0.001	45
SE	0.886 (0.040)	0.812 (0.055)	8.29	0.001	9
MVPA in minutes	50.51 (38.15)	44.33 (33.28)	2.67	0.026	14
Sedentary behaviour	0.660 (0.166)	0.698 (0.164)	3.44	0.007	5

Sample size = 10

*df* degrees of freedom, *TST* total sleep time, *WASO* waking after sleep onset; *SE*: sleep efficiency measured as a proportion on 0–1 scale (see text for details), *MVPA* moderate or vigorous physical activity, *Sedentary Behaviour* is measured as a proportion on 0–1 scale (see text for details), *MAPE* mean average percentage error of the ConD versus ResD (to nearest whole number)

**Table 2** Bias and limits of agreement between a commercial grade (ConD) and research grade device (ResD) for five selected health markers (see plots in Figs. 1, 2, 3, 4)

Variable	Bias	Lower and upper limits of agreement
TST in minutes	+56.41	-45.27, +158.10
WASO in minutes	-29.65	+3.78, -63.10
SE	+0.074	+0.020, +0.126
MVPA in minutes	+6.17	-5.68, +18.03
Sedentary behaviour	-0.038	+0.042, -0.118

Bias is estimated as the mean difference for ConD-ResD values (see text for details); limits of agreement are estimated as  $\pm 1.96 \times \text{s.d. of bias}$

TST total sleep time, WASO waking after sleep onset, SE sleep efficiency measured as a proportion on 0-1 scale (see text for details), MVPA moderate or vigorous physical activity; Sedentary Behaviour is measured as a proportion on 0-1 scale (see text for details)

Feedback from nine participants suggested that whilst the ConD was highly acceptable there was less enthusiasm for the ResD. This appeared to focus on two main issues. First, whilst no individual refused to wear the ResD, five individuals expressed concern that wearing an actiwatch identified them as ‘a patient’. Three of these individuals indicated that they regarded wearing a medical device as potentially stigmatising and said they would be reluctant to use an actiwatch for an extended period. In addition, six individuals reported that they were disappointed that there is no option to review

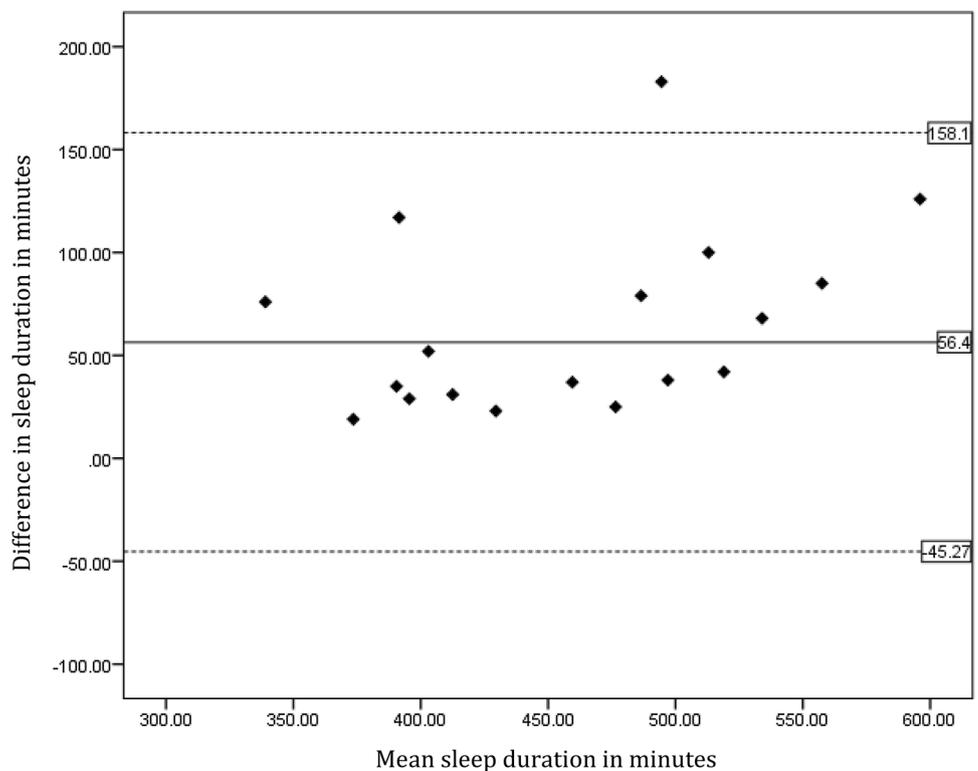
the ResD recordings for themselves and that they preferred the ConD because it allowed them to examine their daytime activity and sleep pattern in real-time using the device display or web-based interface on their phone or tablet.

### Conclusions

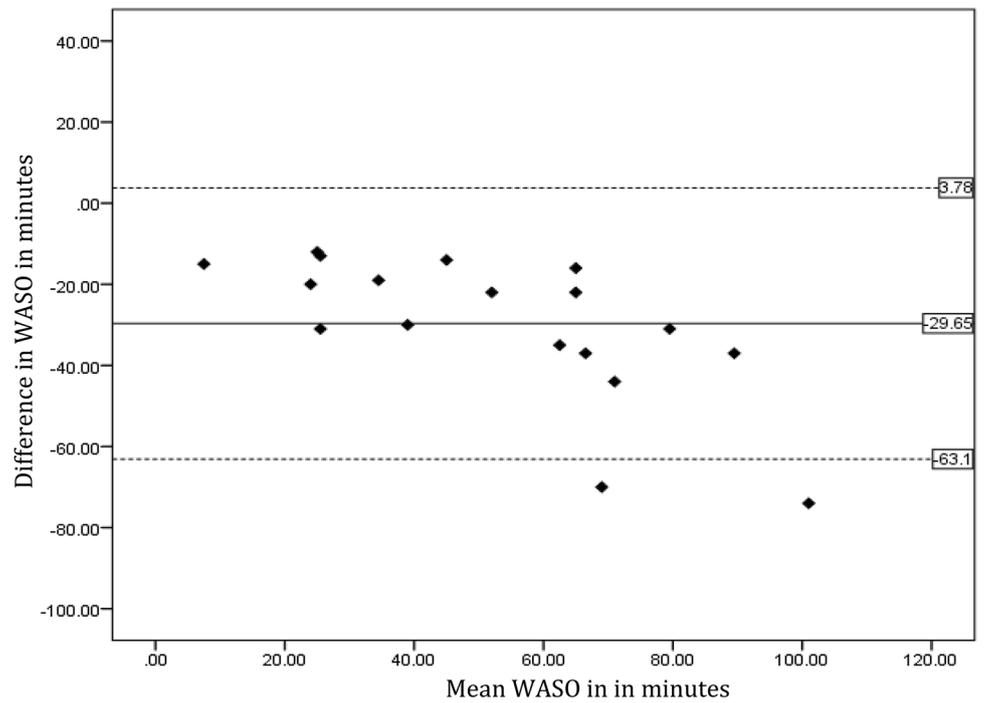
Low levels of daytime activity and disrupted sleep patterns in youth are not only a public health concern, but also are increasingly regarded as important targets for assessment and monitoring in YMH settings [28]. This prompted us to examine whether a consumer grade rest-activity tracker might be used as a substitute for a research grade actigraphic device, and in what circumstances it might be employed. We discuss in the findings and limitations of the study and consider the implications for both research and clinical practice from the perspective of utility and acceptability of the different tools.

Seventy percent (10 of 13) of YMH attendees who consented to participate in this study provided simultaneously recorded data from the ResD and ConD. Using ResD outputs as the criterion values, the findings indicate that the ConD shows only modest levels of accuracy overall (MAPE range from 4 to 45%). As in some previous studies, the use of ConD for monitoring sleep is undermined by the significant overestimation of TST and under-estimation of WASO (and the consequence impact on SE estimation) [20, 23, 29, 30].

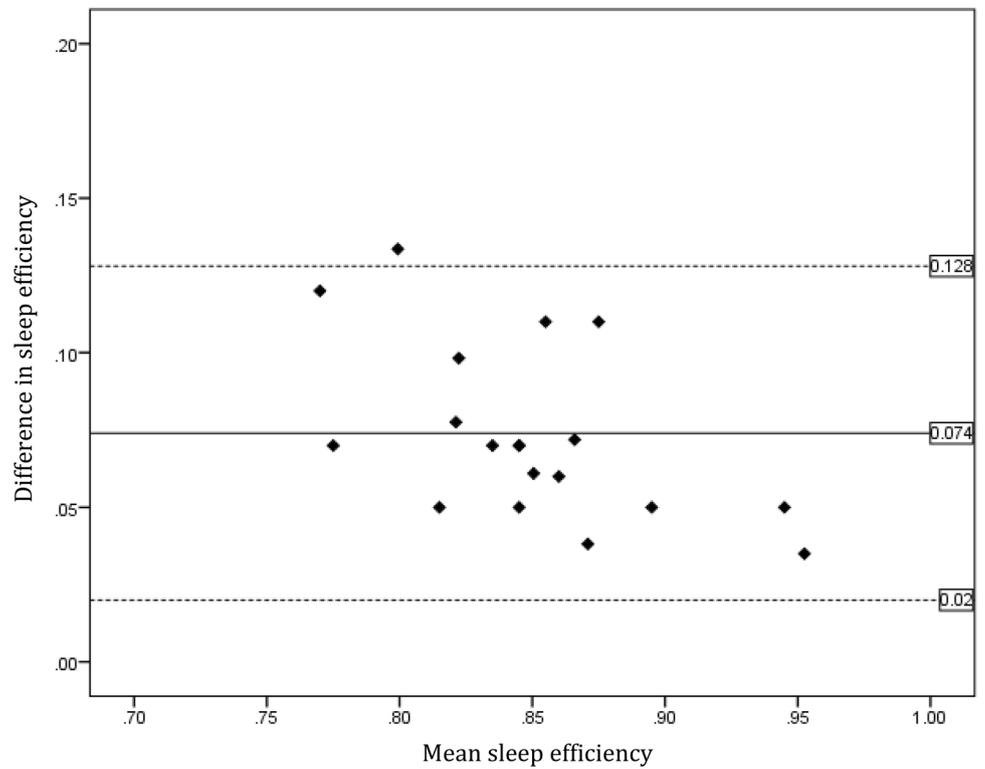
**Fig. 1** Sleep duration in minutes



**Fig. 2** Wake after sleep onset (WASO) in minutes



**Fig. 3** Sleep efficiency

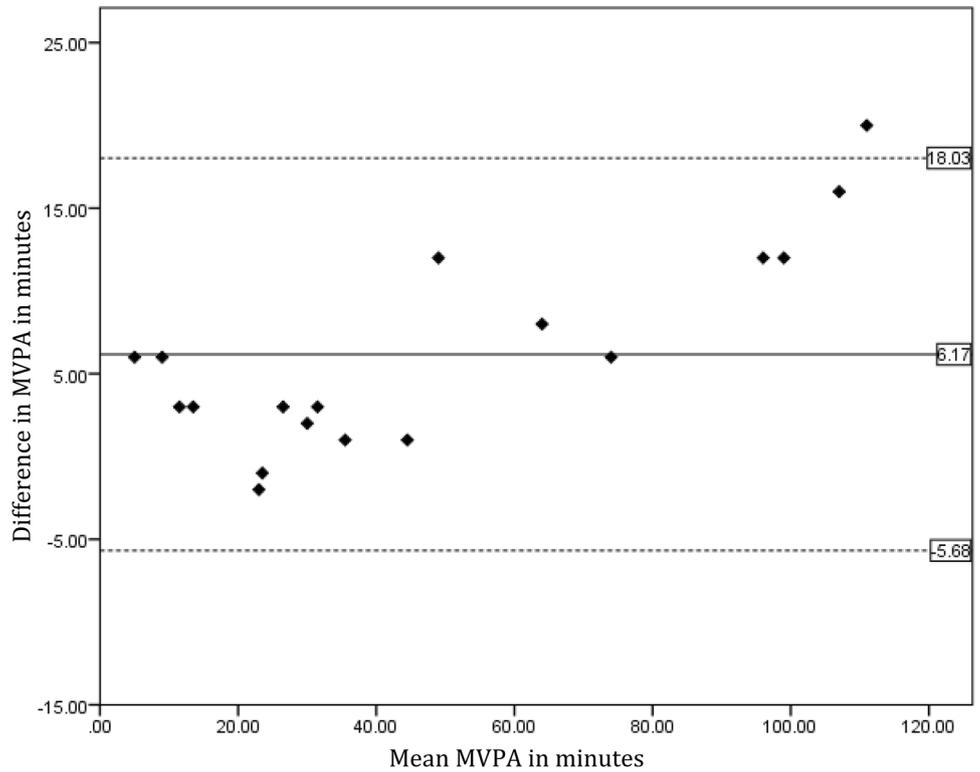


Although MVPA is overestimated by the ConD, the daytime activity metrics appear to be slightly more representative of the ResD values than the sleep metrics [18, 19, 31].

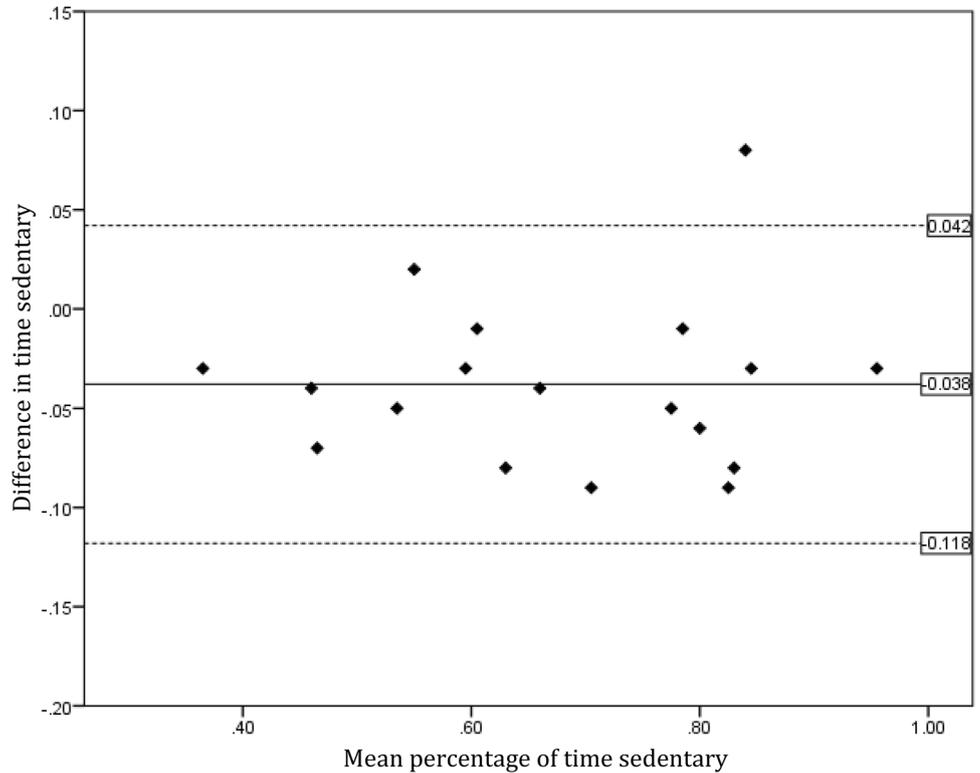
The strength of this study is that metrics and analyses were selected a priori and the sample size was calculated

to allow use of paired *t* tests, MAPE and bias estimates (using Bland–Altman methods). This approach is preferable to correlational analyses as these may conflate apparent agreements between ConD and ResD [32], whilst concealing the magnitude of any errors and/or whether

**Fig. 4** Minutes spent in moderate or vigorous physical activity (MVPA)



**Fig. 5** Percentage of time sedentary



discrepancies are primarily due to over- or under-estimation of values [33]. However, the current study has several limitations. For example, we cannot definitively

state that other rest-activity metrics extracted from ConD or ResD will show the biases reported for the five variables selected, although the indications in the emerging

literature is that the emerging literature is consistent with our findings (and indicates problems with other measures such as sleep latency onset) [18–23, 31]. The metrics we chose were selected because of their widespread use in research and clinical practice, but also because it was easy to extract data for the variables and it was possible to examine them without recourse to sleep or activity diaries (although the ideal would be to use self-report diaries alongside objective data collection, we wished to minimize participant burden). Further, we chose an activity tracker produced by one manufacturer. Whilst this decision is justifiable, we cannot simply extrapolate findings from this ConD to other devices. The pre-study review indicated that the device we chose was the optimal ConD available for the study, but this is a rapidly developing field and testing of new ConD or new models of existing devices will be necessary. In addition, the 7-day recording period was appropriate for comparing the mean values for sleep and activity parameters, but a longer duration is required to extend analyses to, e.g. the study of variability [34]. This may create further issues in collecting data from commercial devices (as batteries need to be recharged every 5 days). Lastly, the sample size was sufficient for the primary analyses, but is insufficient to explore potential confounders of reported findings or predictors of the biases observed.

The data recordings from ConD and ResD are derived from accelerometers and the presumption is that wrist movements can be used as a proxy for monitoring daytime activity patterns, whilst the absence of movement at night equates to sleep. We found that the ConD was insufficiently sensitive to night-time movement; thus, the current study suggests ConD are a relatively poor substitute for ResD for research targeted specifically at sleep patterns. We speculate that this could indicate that, although the ConD contains a triaxial accelerometer system, the calibration of the system for the horizontal plane may be less accurate (the problem of over-estimating TST and under-estimating WASO was reversed but not resolved using the higher sensitivity setting on the ConD). We emphasize that this is a hypothesis that requires testing, but the normal setting produced significantly different values for the sleep metrics compared to the ResD. In contrast, daytime estimates of MVPA and sedentary behaviour (mostly sitting rather than lying down) appeared to be more comparable with the actiwatch recordings. However, from a technological point of view, we do not know for certain if it is the hardware or the software that performs less well in the ConD (or both). As such, the findings warrant further testing in larger studies with, e.g. samples undertaking more detailed, systematic clinical assessments; use of self-report diaries to clarify the nature of rest-activity behaviours; comparisons with other gold standard measures;

using different types of ConD; experimenting with different device sensitivity settings; and longer periods of recording.

One of the reasons for increased interest in objective measures of sleep and activity cycles is their potential use in personalised or precision medicine [35]. At this stage, assuming our findings is confirmed by others, we do not recommend employing the ConD for personalised diagnostics in youth. For instance, using a ConD is unlikely to help to screen for or reliably determine whether an individual is experiencing a clinically meaningful delayed sleep phase; nor is a ConD likely to provide sufficient information to enhance treatment selection. However, ConD might be useful, e.g. to assess within-subject changes in daytime activity or sleep patterns in situations where absolute accuracy is not required. This application could help in clinical practice where intra-individual monitoring may help to ascertain whether a specific intervention or treatment (that has already been selected) is having a positive effect on activity levels or sleep patterns [8, 36]. Further, the availability and acceptability of ConD and the opportunity for real-time self-monitoring may increase their potential utility as an adjunct to repeated prospective assessments of clinical progress. In youth, their preference for using a ConD may increase the likelihood of engagement in monitoring rest-activity patterns and therefore increasing the amount of data available to a clinician or researcher. In addition, its use may enhance collaboration between clinicians and clients.

In conclusion, an advantage of research grade acti-watches is that they measure and store much more data and the algorithms allow evaluation of many more sleep-wake cycle parameters than ConD. A major disadvantage is that actiwatch data are not readily available to the wearer and youth are ambivalent or against their routine use. However, the perceived benefits of ConD in youth may be insufficient in situations where reliable data on a sophisticated set of parameters, as available from ResD, are critical to the decision-making process. Thus, the current study suggests that ResD is preferable for precision diagnostics or attempts to stratify cases into treatment relevant subgroups. For other purposes, there may be a need to consider the ‘trade off’ between quality of data recording versus the likelihood of obtaining the required quantity of sleep and activity data.

**Acknowledgements** JS, HK and IBH are members of the collaborative network called mMARCH (which is involved in international research on activation), which is a workgroup of NIMH, USA that is coordinated by Dr Kathleen Merikangas.

**Author contributions** JS wrote the initial draft of the manuscript. All authors contributed to redrafting and redaction.

**Funding** The study was pilot work for projects undertaken by the researchers (no unique source of funding was obtained specifically for

this project). UK funding (JS) includes the Searle Foundation (competitive award from a private benefactor) and Research for Patient Benefit programme (PB-PG-0609-16166): early identification and intervention in young people at risk of mood disorders. Australian funding (IBH) includes optimising Early Interventions for Young People with Emerging Mood Disorders (APP 1061043) and ‘Testing and delivering early interventions for young people with depression’ (APP 1046899).

**Data Availability** De-identified data from the database can be made available from the corresponding author upon reasonable request.

## Compliance with ethical standards

**Ethics statement** The study complies with the Declaration of Helsinki (the ethical approval and consent procedures are described in the text).

**Conflict of interest** All authors declare they have no links with Fitbit or Fitabase and that the research was undertaken without any input from these organizations. JS is a visiting professor at the Brain and Mind Centre at the University of Sydney. JS has received UK Grant funding from the Medical Research Council (including for projects on actigraphy and bipolar disorders) and from the Research for Patient Benefit programme (PB-PG-0609-16166: early identification and intervention in young people at risk of mood disorders). IBH is a Commissioner in Australia’s National Mental Health Commission; a Member of the Medical Advisory Panel for Medibank; a Board Member of Psychosis Australia Trust. IBH has received honoraria for presentations of his own work at educational seminars supported by a number of non-government organisations and by the pharmaceutical industry (including Servier, Pfizer, AstraZeneca and Eli Lilly). The University of Sydney (Principal Investigator: IH) received funding from Servier for a study of major depression and sleep disturbance in primary care settings. Other relevant funding for IBH in relation to this study includes NHMRC Centre of Research Excellence in Optimising Early Interventions for Young People with Emerging Mood Disorders (APP 1061043) and ‘Testing and delivering early interventions for young people with depression’ (APP 1046899).

**OpenAccess** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, Sadeh A, Spira AP, Taylor DJ. The SBSM guide to actigraphy monitoring: clinical and research applications. *Behav Sleep Med*. 2015;13(Suppl 1):4–38.
2. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol*. 2011;111(2):187–201.
3. Craig R, Mindell J, Hirani V, editors. *Health Survey for England 2008: Physical activity and fitness*. Leeds: NHS Information Centre; 2009. p. 73–101.
4. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev*. 2011;15(4):259–67.
5. Ritter PS, Marx C, Bauer M, Leopold K, Pfennig A. The role of disturbed sleep in the early recognition of bipolar disorder: a systematic review. *Bipolar Disord*. 2011;13(3):227–37.
6. Scott J, Naismith S, Grierson A, Carpenter J, Hermens D, Scott E, Hickie I. Sleep-wake cycle phenotypes in young people with familial and non-familial mood disorders. *Bipolar Disord*. 2016;18(8):642–9.
7. Shou H, Cui L, Hickie I, Lameira D, Lamers F, Zhang J, Crainiceanu C, Zipunnikov V, Merikangas KR. Dysregulation of objectively assessed 24-hour motor activity patterns as a potential marker for bipolar I disorder: results of a community-based family study. *Transl Psychiatry*. 2017;7(8):e1211.
8. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *J Psychiatry Neurosci*. 2006;31(1):13–20.
9. Takaesu Y, Inoue Y, Ono K, Murakoshi A, Futenma K, Komada Y, Inoue T. Circadian rhythm sleep-wake disorders predict shorter time to relapse of mood episodes in euthymic patients with bipolar disorder: a prospective 48-week study. *J Clin Psychiatry*. 2018;79(1).
10. Robillard R, Hermens DF, Naismith SL, White D, Rogers NL, Ip TK, Mullin SJ, Alvares GA, Guastella AJ, Smith KL, Rong Y, Whitwell B, Southan J, Glozier N, Scott EM, Hickie IB. Ambulatory sleep-wake patterns and variability in young people with emerging mental disorders. *J Psychiatry Neurosci*. 2015;40(1):28–37.
11. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, Meier NF, Dixon PM. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc*. 2016;48(1):151–8.
12. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *Int J Environ Res Public Health*. 2009;6(2):492–525.
13. Cellini N, McDevitt E, Mednick S, Buman M. Free-living cross-comparison of two wearable monitors for sleep and physical activity in healthy young adults. *Physiol Behav*. 2016;157:79–86.
14. Lee HA, Lee HJ, Moon JH, Lee T, Kim MG, In H, Cho CH, Kim L. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. *Psychiatry Investig*. 2017;14(2):179–85.
15. Case M, Burwick H, Volpp K, Patel M. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA*. 2015;313(6):625–6.
16. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):39–49.
17. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? *Med Sci Sports Exerc*. 2000;32(2):426–31.
18. Evenson K, Goto M, Furberg R. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12:159.
19. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act*. 2015;12:42.
20. Kooiman T, Dontje M, Sprenger S, Krijnen W, van der Schans C, de Groot M. Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil*. 2015;7:24.
21. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. *Health Psychol Open*. 2016;3(2):2055102916679012.
22. de Zambotti M, Baker F, Willoughby A, Godino J, Wing D, Patrick K, Colrain I. Measures of sleep and cardiac functioning

- during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav.* 2016;158(1):143–9.
23. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J Med Eng Technol.* 2017;41(3):208–15.
  24. Ekblom O, Nyberg G, Bak E, Ekelund U, Marcus C. Validity and comparability of a wrist-worn accelerometer in children. *J Phys Act Health.* 2012;9(3):389–93.
  25. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. *Prog Cardiovasc Dis.* 2015;57(4):387–95.
  26. 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.
  27. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327:307–10.
  28. Zhang J, Paksarian D, Lamers F, Hickie IB, He J, Merikangas KR. Sleep patterns and mental health correlates in US adolescents. *J Pediatr.* 2017;182:137–143.
  29. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. *Sleep.* 2007;30:1362.
  30. Lee PH, Suen LK. The convergent validity of Actiwatch 2 and ActiGraph Link accelerometers in measuring total sleeping period, wake after sleep onset, and sleep efficiency in free-living condition. *Sleep Breath.* 2017;21(1):209–15.
  31. Baroni A, Bruzzese JM, Di Bartolo CA, Shatkin JP. Fitbit Flex: an unreliable device for longitudinal sleep measures in a non-clinical population. *Sleep Breath.* 2016;20(2):853–4.
  32. Mooses K, Oja M, Reisberg S, Vilo J, Kull M. Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study. *BMC Public Health.* 2018;18(1):858.
  33. Carstensen B, Simpson J, Gurrin LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. *Int J Biostat.* 2008;4(1):Article 16.
  34. Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, Merikangas K, Hickie IB. Activation in bipolar disorders: a systematic review. *JAMA Psychiatry.* 2017;74(2):189–96.
  35. Scott J, Etain B, Bellivier F. Can an integrated science approach to precision medicine research improve lithium treatment in bipolar disorders? *Front Psychiatry* (2018) (**in press**).
  36. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. *BMC Public Health.* 2015;15:585.

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