Validity and reliability of the Fitbit Zip as a measure of preschool children's step count

Catherine A Sharp,[1] Kelly A Mackintosh,[2] Mihela Erjavec,[1] Duncan M Pascoe,[1] Pauline J Horne[1]

ABSTRACT

Objectives Validation of physical activity measurement tools is essential to determine the relationship between physical activity and health in preschool children, but research to date has not focused on this priority. The aims of this study were to ascertain inter-rater reliability of observer step count, and interdevice reliability and validity of Fitbit Zip accelerometer step counts in preschool children.

Methods Fifty-six children aged 3–4 years (29 girls) recruited from 10 nurseries in North Wales, UK, wore two Fitbit Zip accelerometers while performing a timed walking task in their childcare settings. Accelerometers were worn in secure pockets inside a custom-made tabard. Video recordings enabled two observers to independently code the number of steps performed in 3 min by each child during the walking task. Intraclass correlations (ICCs), concordance correlation coefficients, Bland-Altman plots and absolute per cent error were calculated to assess the reliability and validity of the consumer-grade device.

Results An excellent ICC was found between the two observer codings (ICC=1.00) and the two Fitbit Zips (ICC=0.91). Concordance between the Fitbit Zips and observer counts was also high (r=0.77), with an acceptable absolute per cent error (6%–7%). Bland-Altman analyses identified a bias for Fitbit 1 of 22.8±19.1 steps with limits of agreement between −14.7 and 60.2 steps, and a bias for Fitbit 2 of 25.2±23.2 steps with limits of agreement between −20.2 and 70.5 steps.

Conclusions Fitbit Zip accelerometers are a reliable and valid method of recording preschool children's step count in a childcare setting.

What are the new findings

- Placing a Fitbit Zip in the pocket of a novel and low-maintenance tabard is a valid and acceptable method for a young child to wear the device.
- Fitbit Zip is a cheaper alternative to expensive research-grade devices, making accelerometers more accessible for large-scale trials on physical activity with preschoolers.
- Fitbit Zip is a valid and reliable activity monitor to record preschool children's step counts in their natural settings.

INTRODUCTION

With more than 42 million preschool children classified as overweight or obese globally,[1] early intervention to increase daily physical activity levels, and promote positive lifestyle behaviours that will track into adulthood, is a public health priority. It has been estimated that physical inactivity causes 6%–10% of major non-communicable diseases worldwide, such as coronary heart disease, breast and colon cancers, and type 2 diabetes.[2] In the UK, children aged 3–4 years are entitled to 15 hours per week of free early education,[3] providing an important opportunity for them to work towards their recommended target of 180 min of daily activity set by the Department of Health,[4] a target equated empirically to 6000 steps per day.[5] However, a major public health concern is that children in childcare are not sufficiently active[6]; daily time spent engaging in sedentary behaviour was greater in children who attended childcare for 6 hours as compared with those who attended for 3 hours.[7] Indeed, our own extensive observations of children following the Early Years Foundation Phase Framework (data not provided) confirm that young children in their childcare sessions tend to either walk around or engage in sedentary activities while seated at a table or on the floor. In order to measure young children's physical activity in the childcare setting, it is important to identify reliable, accessible and wear-compliant measurement tools.[8][9] While activity monitors have been found to provide valid measures of primary school children's step counts, researchers cannot assume that they will also provide valid step count measures in preschool children.[10][11]

Given that Public Health targets for children's daily physical activity levels tend to be

[1]School of Psychology, Bangor University, Bangor, Gwynedd, UK
[2]Research Centre in Applied Sports, Technology, Exercise and Medicine, Swansea University, Swansea, UK

Correspondence to
Dr Catherine A Sharp; c.sharp@bangor.ac.uk

expressed in terms of steps performed over a specified number of minutes or hours per day, and typically specify walking,[12] step count is regularly reported when assessing physical activity levels in everyday settings. To date, validated devices most often used to measure preschool children's step counts are research-grade accelerometers (eg, Actigraph[13]) and pedometers (eg, Omron[14]). However, such devices have inherent limitations, such as cost, intensity debates, data interpretation and compliance issues, particularly in the case of young children.[15 16] Recently, a number of widely available, consumer-grade accelerometers have been trialled as potentially reliable and low-cost measures of physical activity for use in research studies,[17] with the 'Fitbit' brand identified as the most popular range.[17 18] One function of Fitbit devices is the measurement of total step count, as well as steps performed minute by minute over the recording period. When compared with nine similar activity trackers, the Fitbit Zip was found to provide the most valid measure of step count.[19] However, previously, the Fitbit Zip has only been validated for healthy adults[17 20] and older adults.[21]

The present study assessed Fitbit Zip interdevice reliability, inter-rater reliability of observer step count and validity of step counts in preschool children within a childcare setting. Findings from the present study informed subsequent use of the device in a controlled evaluation of a behaviour change activity intervention in childcare settings.[22]

METHOD

Participants

Opportunity sampling at 10 nurseries in North Wales, UK, enabled recruitment of 66 preschool children (33 girls, 3.7±0.6 years) to participate in a single-session, timed walking task. Informed parental consent was obtained for the children's participation. The data from 10 children were excluded prior to coding: 5 because they did not perform the task as requested and 5 due to technical errors in their video recordings. As this is the first validation of a consumer-level activity monitor with preschool children, previous data were not available to inform the sample size calculation for this target population. Therefore, in line with conservative findings (correlation of 0.5) used in a previous study comparing a Fitbit accelerometer and observer step count with adults,[23] an a priori sample size was calculated in G*Power as 42 participants (α=0.05, and a power of 0.95). The School of Psychology Ethics and Research Governance Committee at Bangor University granted ethical consent for the study (ethics application number: 2013-11864).

Measurement

The Fitbit Zip is a small (width 2.8 cm×height 3.6 cm×depth 1.0 cm), lightweight (8 g), inexpensive (£49.99) and water-resistant commercial activity monitor, which contains a microelectromechanical tri-axial accelerometer and uses proprietary algorithms to calculate step counts recorded on the device. The majority of accelerometers and pedometers are typically secured to the right hip using an elastic belt and require adult (researcher/parent) input to ensure continued compliance and correct placement.[24] Based on the manufacturer's claim that the Fitbit Zip can be 'worn on or very close to the body', including external pieces of clothing,[25] each child wore two devices ('Fitbit 1' and 'Fitbit 2') securely positioned inside a custom-made, close-fitting cotton tabard with elasticated sides, which was designed to cover the children's own everyday clothing (see figure 1). The use of elasticated sides provided a close and comfortable fit, while simultaneously ensuring that the tabard with enclosed Fitbits always moved in exactly the same directions as the child's body. The function of the tabard was to combat issues of wear compliance identified with other accelerometers.[16] The pockets on the inner face of the tabard were positioned one above the other over the child's right hip (Fitbit 1 placed in the upper pocket).

Protocol

Following consent, children were invited one at a time to take part in a 'walking adventure', while wearing the tabard with two Fitbit Zips in situ. The walking task was first demonstrated to the child, who was then invited to perform a practice walk with the researcher before performing the task independently. Participants were asked to walk back and forth between 'point A' and 'point B' for approximately 5 min in an open space in each nursery. Their performance was recorded using a video camera (Sony HDR-GW77E CX115), which was positioned in line with each child's direction of travel during the walking task. Episodes from 'Thomas the Tank Engine' cartoon were displayed on an audiovisual device (Apple iPad) at point B, as entertainment for the children during the task. Verbal encouragement was given throughout. Following the task, recorded data were transferred wirelessly using Bluetooth technology to Fitbit's standard application programming interface. Minute-by-minute data were subsequently extracted through a third-party company (whatAdata, Swansea, UK).

Observer coding of the children's steps from the video footage was conducted independently by two individuals, an expert coder to provide the criterion measure and a trainee coder to enable assessment of inter-rater reliability. While similar validation studies using direct observation have not published their coding framework, the present study defined a step as 'Lifting and setting down one's foot or one foot after the other in order to walk somewhere or move to a new position'.[26] The child's entire foot was required to leave the floor completely to be coded as a step. If both feet left the ground simultaneously during locomotion (ie, the child 'jumped') this counted as two steps as

Figure 1 Custom-made tabard with inside pockets securing the two Fitbit Zips in position (contact first author for more details on tabard construction). Consent was obtained for the publication of the child's photograph.

children typically landed on one foot first and then the other to regain their balance. If a child slid their feet along the ground or the movement was so small that the coders could not identify a step-like movement, a 'no-step' was recorded.

To help train consistency, the two coders initially took part in practice sessions using the coding framework. The footage was coded at 35% full speed. The coders used a hand-held tally counter, and to help maintain their attention on the coding task, they articulated orally their recording of each valid step. This method enabled the coders to maintain their visual focus on the footage displayed on the screen. To identify each minute, the footage was time stamped with hours, minutes and seconds. Where a step occurred during the crossover of a minute, the step was attributed to the minute in which contact with the floor was re-established. Step count was first coded independently by the expert coder and the trainee coder. Following identification of several sections of the recordings where expert counts (ie, criterion) differed from trainee counts by more than six steps, the trainee

coder expressed uncertainty as to how a child's 'jump' should be coded. After a reminder of the operational definition, the trainee independently coded the relevant sections once again. This instance of 'observer drift' will inform training of new coders in subsequent research studies.

Three minutes of performance on the walking task were coded for comparison with the Fitbit Zips. Where a child had more than 3 min of data, a random number generator dictated which minutes within the 4 or 5 min to include. This walking task time sample per participant compares well with the 2 min used with adults.[21] For each child, a 'total number of steps' variable was calculated for each of the four measures (two observer counts; two Fitbit Zips).

Statistical analysis

Data analyses were conducted using PASW Statistics V.22 (SPSS). No statistical outliers were identified. Intraclass correlations (ICCs) were calculated for each measure to assess reliability using recommended procedures.[27] Reliability of the trainee's counts was assessed using a two-way mixed-effects model with absolute agreement applied. Interdevice reliability was assessed using a two-way random-effects model again with absolute agreement applied. Concordance correlation coefficients (CCCs) were calculated to evaluate consistency between the observer's step counts for each participant.[28] To interpret the findings, the following cut-off criteria were used: an ICC or CCC of 0.75 and above is classed as 'excellent'; 0.60–0.74 as 'good'; 0.40–0.59 as 'fair' and 0.39 and below as 'poor'. Bland-Altman plots were used to investigate agreement between the two measurements.[29 30] To enable comparison between the devices, absolute per cent error was calculated and interpreted by the standard of ±10% error for free-living conditions.[21 31] The equation was as follows: ((Fitbit output−Observer count)/ Observer count)×100.

RESULTS

The final sample included 56 children (29 girls; mean±SD: 3.69±0.58 years). Excellent intercoder agreement was achieved between the expert coder ($\overline{x}$=367.68± 40.89) and the trainee coder ($\overline{x}$=369.09±41.00; ICC=1.00; 95% CI 0.99 to 1.00). Also, excellent interdevice agreement was found between Fitbit 1 ($\overline{x}$=344.91±41.11) and Fitbit 2 ($\overline{x}$=342.52±52.91; ICC=0.91; 95% CI 0.85 to 0.95). Good concordance was found between the expert coder and both Fitbit 1 (r=0.77; 95% CI 0.66 to 0.85) and Fitbit 2 (r=0.77; 95% CI 0.67 to 0.84). The Bland-Altman plots compared observer counts with Fitbit 1 (figure 2A) and Fitbit 2 (figure 2B). Specifically, analyses identified a bias of 22.8±19.1 steps and 25.2±23.2 steps, with the limits of agreement interval being −14.7 to 60.2 steps and −20.2 to 70.5 steps, for Fitbits 1 and 2, respectively.
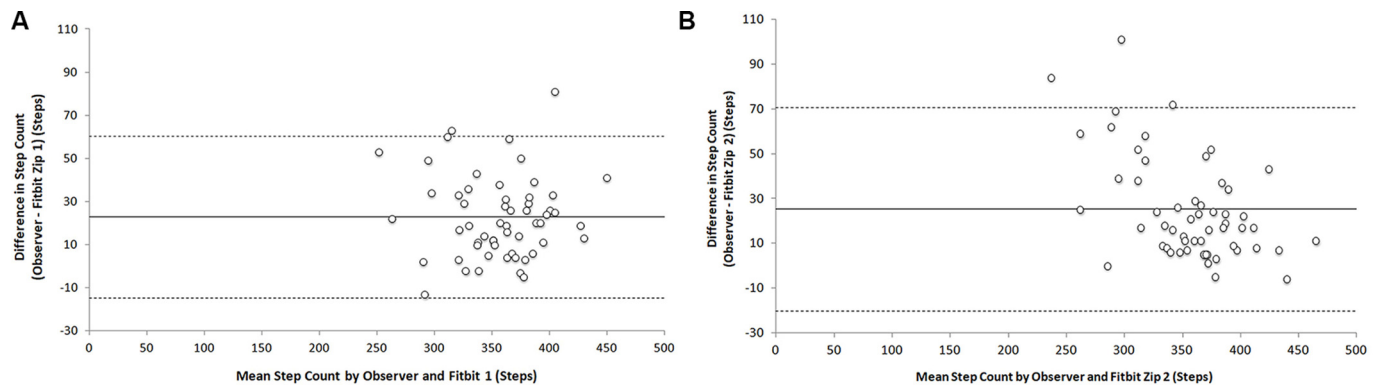
Figure 2 Bland-Altman plots illustrating the relationship between the expert coder and (A) Fitbit 1 and (B) Fitbit 2. Solid line represents the mean difference between the two measures, and dashed lines represent limits of agreement (±1.96 SD).

The absolute per cent error for Fitbit 1 was 6.44% (SE=0.66) and for Fitbit 2 was 7.27% (SE=0.94). The frequencies of overcounting (>10%), exact counting (±10%) and undercounting (≤10%) show that the majority of Fitbit 1 and Fitbit 2 step counts were in the exact-counting range (80.40% and 76.70%, respectively), with the remainder in the undercounting range; no devices overcounted by more than 10% of the children's steps.

DISCUSSION

The main aim of this study was to assess the interdevice reliability and validity of the Fitbit Zip as a measure of preschool children's step counts while performing a walking task in their nursery setting. Our results show excellent interdevice reliability with a high ICC value. Comparisons between Fitbit Zip step counts and expert step counts (criterion measure) found good agreement for both Fitbit Zip placements. Bland-Altman analyses identified an average undercount in the number of steps recorded by Fitbit 1 and Fitbit 2 (22.8 and 25.2 steps, respectively). However, in line with other validation studies, the majority of data points fall within the 95% limits of agreement. Additionally, the low absolute discrepancy between the measures (±10% error) obtained in a non-laboratory setting is consistent with findings in other target populations,[21][31] suggesting that the Fitbit Zip is also a valid and reliable measure of preschool children's step counts in a nursery setting.

Accelerometers are an objective method of assessing preschool children's step count[24] and are typically validated worn on the hip or wrist. When identifying a measurement tool for the present study, practical implications[11] and the developmental stage of preschool children[16] were considered. Low acceptability of devices worn on elasticated belts and the consequent risks of damage or loss of the device led to the development of a novel and low-maintenance means for preschool children to engage freely with their daily environment while wearing activity monitors. It can therefore be postulated that wearing a Fitbit Zip inside the tabard does not adversely impact on the accuracy of the device,

indicating their potential use as an integrative research tool. Indeed, in a subsequent study conducted by the first author, participants aged 3–4 years enjoyed wearing the tabards throughout their entire daily childcare sessions[22] and the nursery staff were happy to place the tabards on the children with monitors in situ at the start of their sessions and remove them at the end.[32] The findings provide support for Fitbit's[25] claims that the Fitbit Zip can be worn in or on external clothing.

The present study is associated with numerous strengths. Conducting the study in a real-world nursery environment increased its ecological validity, although it should be acknowledged that a greater element of error may be expected in comparison to laboratory-based studies.[23][33] The procedures employed are consistent with recent validation methodologies in adults[21] and young children,[12] using direct observation as the criterion measure. Furthermore, the present study also followed the analysis strategy of a recent validation study and related guidelines.[23][27][28] Despite the exclusion of 10 children, the power sample size was maintained. While a 7.5% attrition rate (five children), attributable to failure to complete the task, is very low as compared with developmental studies with children of a similar age,[34] it highlights the importance of over-recruiting participants, particularly when working with preschool children. Some children of this age can be highly distractible and therefore less likely to complete the task, in comparison to older children. Comparison of the outcomes found here for the Fitbit Zip with those for alternative devices has proved difficult because published studies aiming to validate those tools in preschool children have typically reported only correlation coefficients,[15] which can be misleading.[30] As others have argued, measures of correlation and measures of agreement do not evaluate the same construct.[11][15][30] These studies have also not reported absolute per cent error. The use of the gold-standard criterion measure of direct observation is a key strength of this study. A previous study in older adults employing a 2 min walking task compared direct observation, Fitbit Zips and ActiGraphs, and found that

Fitbit devices had a higher agreement with step counts recorded by direct observation than with those recorded by ActiGraphs.[21]

Despite the strengths of the paper, limitations must be acknowledged. The spatial placement limits of this interdevice reliability could be investigated further by systematically varying the locations of the devices worn in the pockets of the children's tabards. For example, devices could be worn over the right and left hips, at both the front and back, and at the child's chest level. In the current study, despite Fitbit 1 being placed only 1 cm above Fitbit 2, a greater bias for undercounting steps taken by the children was seen for the lower Fitbit 2 location. Future studies should evaluate interinstrument and intrainstrument reliability of Fitbit accelerometers to better determine whether reliability issues may have contributed to slight differences in the data captured in this study. Counterbalancing the placement of the two accelerometers across participants would have helped differentiate the source of bias (ie, the particular device from the particular wear location). For example, the study could be repeated with both Fitbit Zips side by side in the same pocket, with left and right placement randomised across children to provide a more refined, location-independent measure of interdevice reliability.

CONCLUSIONS

This study, to our knowledge, is the first to validate a consumer-grade activity monitor in a preschool education setting showing that the device is an accurate tool to assess preschool children's habitual step count. The Fitbit Zips were also found to have strong interdevice reliability and appear to provide a valid and cheaper alternative to research-grade activity devices, well suited to measuring steps performed by young children in relation to current Public Health targets. The Fitbit Zip would also be an excellent tool for measuring preschool children's habitual physical activity in terms of step counts within and across cultures, which should be further explored.

REFERENCES

1. World Health Organisation. 2013 childhood overweight and obesity. http://www.who.int/dietphysicalactivity/childhood/en/ (accessed 04 Apr 2016).
2. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012;380:219–29.
3. British Government. Help paying for childcare. https://www.gov.uk/help-with-childcare-costs/free-childcare-and-education-for-2-to-4-year-olds (accessed 12 Mar 2017).
4. Department of Health. *Start active, stay active: a report on physical activity for health from the four home countries chief medical officers*, 2011. http://www.bhfactive.org.uk/userfiles/Documents/startactivestayactive.pdf (accessed 04 Apr 2016).
5. Gabel L, Proudfoot NA, Obeid J, et al. Step count targets corresponding to new physical activity guidelines for the early years. *Med Sci Sports Exerc* 2013;45:314–8.
6. Reilly JJ. Low levels of objectively measured physical activity in preschoolers in child care. *Med Sci Sports Exerc* 2010;42:502–7.
7. O'Dwyer MV, Fairclough SJ, Ridgers ND, et al. Effect of a school-based active play intervention on sedentary time and physical activity in preschool children. *Health Educ Res* 2013;28:931–42.
8. Hnatiuk JA, Salmon J, Hinkley T, et al. A review of preschool children's physical activity and sedentary time using objective measures. *Am J Prev Med* 2014;47:487–97.
9. Timmons BW, Leblanc AG, Carson V, et al. Systematic review of physical activity and health in the early years (aged 0-4 years). *Appl Physiol Nutr Metab* 2012;37:773–92.
10. Cliff DP, Reilly JJ, Okely AD. Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0-5 years. *J Sci Med Sport* 2009;12:557–67.
11. Kelly LA, Reilly JJ, Fairweather SC, et al. Comparison of two accelerometers for assessment of physical activity in preschool children. *Pediatr Exerc Sci* 2004;16:324–33.
12. Tudor-Locke C, Lutes L. Why do pedometers work? A reflection upon the factors related to successfully increasing physical activity. *Sports Med* 2009;39:981–93.
13. Pate RR, Almeida MJ, McIver KL, et al. Validation and calibration of an accelerometer in preschool children. *Obesity* 2006;14:2000–6.
14. De Craemer M, De Decker E, Santos-Lozano A, et al. Validity of the omron pedometer and the actigraph step count function in preschoolers. *J Sci Med Sport* 2015;18:289–93.
15. Oliver M, Schofield GM, Kolt GS. Physical activity in preschoolers: understanding prevalence and measurement issues. *Sports Med* 2007;37:1045–70.
16. Costa S, Barber SE, Cameron N, et al. The objective measurement of physical activity and sedentary behaviour in 2-3 year olds and their parents: a cross-sectional feasibility study in the bi-ethnic Born in Bradford cohort. *BMC Public Health* 2015;15:1109.
17. Ferguson T, Rowlands AV, Olds T, et al. 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;12:42–51.
18. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015;12:159–81.
19. Kooiman TJ, Dontje ML, Sprenger SR, et al. Reliability and validity of ten consumer activity trackers. *BMC Sports Sci Med Rehabil* 2015;7:25–36.
20. Tully MA, McBride C, Heron L, et al. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Res Notes* 2014;7:952–7.

21. Paul SS, Tiedemann A, Hassett LM, *et al*. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. *BMJ Open Sport Exerc Med* 2015;1:e000013–15.

22. Sharp CA, Horne PJ, Erjavec M, *et al*. Let's have some adventures: increasing preschool children's physical activity levels with a role-modelling intervention. *Poster presented at 8th ABAI International Conference* 2015.

23. Takacs J, Pollock CL, Guenther JR, *et al*. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport* 2014;17:496–500.

24. Pate RR, O'Neill JR, Mitchell J. Measurement of physical activity in preschool children. *Med Sci Sports Exerc* 2010;42:508–12.

25. Fitbit. *How do I wear my tracker*, 2015. https://help.fitbit.com/articles/en_US/Help_article/How-do-I-wear-my-tracker (accessed 05 May 2015).

26. Oxford Dictionary. *Step*, 2015. http://www.oxforddictionaries.com/definition/english/step (accessed 02 May 2015).

27. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol* 2012;8:23–34.

28. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255–68.

29. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.

30. Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015;25:141–51.

31. Tudor-Locke C, Sisson SB, Lee SM, *et al*. Evaluation of quality of commercial pedometers. 2006;97:S10–15.

32. Sharp CA, Horne PJ, Lowe CF, *et al. Evaluation of the super dynamic food dudes intervention for 3-4 year old children at school. Paper presented at Experimental Analysis of Behaviour Group.* London, England, 2017.

33. Diaz KM, Krupka DJ, Chang MJ, *et al*. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. *BMC Res Notes* 2016;9:213–22.

34. Howie EK, Straker LM. Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: a brief methodological review. *J Sci Med Sport* 2016;19:830–6.