Convergent Validity of the Fitbit Charge 2 to Measure Sedentary Behavior and Physical Activity in Overweight and Obese Adults

Joanne A. McVeigh

Curtin University and University of Witwatersrand

Jennifer Ellis, Caitlin Ross, Kim Tang,
Phoebe Wan, and Rhiannon E. Halse

Curtin University

Satvinder Singh Dhaliwal

Curtin University and National University of Singapore

Deborah A. Kerr and Leon Straker

Curtin University

Activity trackers provide real-time sedentary behavior (SB) and physical activity (PA) data enabling feedback to support behavior change. The validity of activity trackers in an obese population in a free-living environment is largely unknown. This study determined the convergent validity of the Fitbit Charge 2 in measuring SB and PA in overweight adults. The participants ($n = 59$; $M \pm SD$: age = 48 ± 11 years; body mass index = 34 ± 4 kg/m²) concurrently wore a Charge 2 and ActiGraph GT3X+ accelerometer for 8 days. The same waking wear periods were analyzed, and standard cut points for GT3X+ and proprietary algorithms for the Charge 2, together with a daily step count, were used. Associations between outputs, mean difference (MD) and limits of agreement (LOA), and relative differences were assessed. There was substantial association between devices (intraclass correlation coefficients from .504, 95% confidence interval [.287, .672] for SB, to .925, 95% confidence interval [.877, .955] for step count). In comparison to the GT3X+, the Charge 2 overestimated SB (MD = 37, LOA = -129 to 204 min/day), moderate to vigorous PA (MD = 15, LOA = -49 to 79 min/day), and steps (MD = 1,813, LOA = -1,066 to 4,691 steps/day), and underestimated light PA (MD = -32, LOA = -123 to 58 min/day). The Charge 2 may be a useful tool for self-monitoring of SB and PA in an overweight population, as mostly good agreement was demonstrated with the GT3X+. However, there were mean and relative differences, and the implications of these need to be considered for overweight adult populations who are already at risk of being highly sedentary and insufficiently active.

Keywords: accelerometer, overweight adults, step count, wearable activity tracker

The increasing prevalence of obesity is a major public concern. Obesity is associated with detrimental health outcomes and places a large burden on a society's economy (Hayes, Lung, Bauman, & Howard, 2017). A recent review, which examined the economic burden of obesity internationally, showed that, on average, obese people have medical costs that are 32% higher compared with those of normal weight (Yusefzadeh, Rahimi, & Rashidi, 2019). Healthy lifestyle changes have been universally advocated, yet the prevalence of overweight and obesity continues to increase (Roberto et al., 2015).

Wearable activity trackers provide real-time data to track sedentary behavior (SB) and physical activity (PA) and provide feedback for participants to support behavior change. As such, these devices have the potential for use in obesity interventions, as a motivational tool to encourage autonomy, and through feedback on performance to guide participants toward a more active lifestyle

(Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015). Since their commercial introduction, activity trackers have grown in popularity, likely as a result of their increasing affordability, sophistication, and smartphone integration (Evenson, Goto, & Furberg, 2015; Feehan et al., 2018). Estimates of adults meeting PA guidelines vary according to the way that the data are measured and the way people report their activity data (Zenko, Willis, & White, 2019). PA guidelines (including step-count-based ones) are the same for healthy adults, regardless of whether they carry excess body weight. In field-based settings, activity trackers may be a useful bridge for helping to overcome reporting errors by measuring the actual motion of the body rather than participant perceptions of their PA levels and may provide a more accurate measure of these behaviors (Hickey & Freedson, 2016), but an awareness of the caveats of activity trackers is important.

The majority of consumer activity tracker validation studies have been conducted in populations with a body mass index (BMI) of between 18.5 and 25 (Tully, McBride, Heron, & Hunter, 2014), laboratory-based (Sushames, Edwards, Thompson, McDermott, & Gebel, 2016), or only evaluated for specific activity types, such as walking (Huang, Xu, Yu, & Shull, 2016). Only a few studies have examined the validity of activity trackers in free-living settings, with findings indicating variability in measurement parameters between devices and between participants (Evenson et al., 2015; Hargens, Deyarmin, Snyder, Mihalik, & Sharpe, 2017; Júdice, Santos, Hamilton, Sardinha, & Silva, 2015; Mikkelsen et al., 2020;

McVeigh, Ellis, Ross, Tang, and Wan are with the School of Occupational Therapy, Speech Therapy and Social Work, Curtin University, Perth, Western Australia, Australia. McVeigh is also with the Movement Physiology Laboratory, School of Physiology, University of Witwatersrand, Johannesburg, South Africa. Halse, Dhaliwal, and Kerr are with the School of Public Health, Curtin University, Perth, Western Australia, Australia. Dhaliwal is also with the Duke-NUS Medical School, National University of Singapore, Singapore. Straker is with the School of Physiotherapy and Exercise Science, Curtin University, Perth, Western Australia, Australia. McVeigh (Joanne.McVeigh@curtin.edu.au) is corresponding author.

Reid et al., 2017; Toth et al., 2018). Free-living settings are more likely to reflect the usual pattern of daily activities; thus, examining the validity of activity trackers under free-living settings is important (Chu et al., 2017). To successfully support healthier behaviors, activity trackers used by consumers should provide accurate and reliable data to guide lifestyle choices and help consumers know whether they are meeting activity targets, especially if activity data are used to support tailored interventions (Benedetto et al., 2018).

Before activity trackers are applied more broadly into the clinical health field, more research is needed to evaluate their performance accuracy in the free-living environment and specific clinical populations at highest risk for obesity and physical inactivity (Hickey & Freedson, 2016; Yoost, Gerlach, Sinning, & Cyphert, 2018). There is conflicting evidence on whether measurements made by research-grade monitors (such as the ActiGraph GT3X+, ActiGraph, Pensacola, FL) are affected by body weight status, with some studies suggesting they are not affected (Feito, Bassett, Thompson, & Tyo, 2012; Feito, Bassett, Tyo, & Thompson, 2011), while another study found that BMI and abdominal volume did affect the accuracy of the accelerometer step count algorithm, resulting in an underestimation of step count for individuals with a BMI ≥ 30 (Belanger, Kestens, Gilbert, Tremblay, & Mathieu, 2014). The validity of consumer-grade activity trackers has only been investigated in a small sample of overweight individuals in one other study reporting anomalous results (Mikkelsen et al., 2020) and deserves further attention, especially considering that an over- or underestimation of PA behaviors has important consequences for the clinical management of at-risk populations.

Most validation studies of activity trackers have examined Fitbit trackers, due to their market dominance and, hence, population use. This brand has sold more than 63 million products globally in the past 10 years, accounting for approximately 20% of the market share (Feehan et al., 2018). A systematic review concluded that step count measurement (from multiple Fitbit activity trackers) demonstrated moderate accuracy for adults with no mobility limitations when compared with research-grade accelerometers, but that activity trackers are unlikely to record energy expenditure accurately and are likely to overestimate PA, particularly with devices that are worn on the wrist (Feehan et al., 2018).

The only other available study that has compared the Fitbit Charge 2 (Fitbit Inc., San Francisco, CA) with a research-grade accelerometer (ActiGraph GT3X+) in adults with a BMI ≥ 25 showed that the Charge 2 overestimated the step count, light-intensity PA, moderate to vigorous PA (MVPA), and underestimated sedentary time (Mikkelsen et al., 2020). In community-dwelling female adults, the Fitbit One and Fitbit Flex overestimated the time spent in light PA and MVPA, when compared with a waist-worn GT3X+ (Reid et al., 2017). Recently, the Fitbit Flex and GT3X+ were shown to be statistically equivalent to one another (in a convenience sample of adults) in assessing SB, but estimates of time spent in MVPA were significantly overestimated by the Fitbit Flex (Redenius, Kim, & Byun, 2019). In a study of older adults with knee osteoarthritis, the Charge 2 overestimated the step count and daily sedentary time by 39% and 37%, respectively, and underestimated the daily MVPA by 50% compared with the GT3X+ (Collins, Yang, Trentadue, Gong, & Losina, 2019).

This study aimed to evaluate the convergent validity of the wrist-worn Charge 2 activity tracker by comparing output with a hip-worn GT3X+ accelerometer for (a) time spent in SB, (b) time spent in light PA, (c) time spent in MVPA, and (d) step count in community-dwelling adults with a BMI ≥ 25 .

Materials and Methods

Study Design and Participants

The data presented in the current study were from participants enrolled in a larger 1-year randomized controlled trial, the “Tailored Diet and Activity study” (ToDay). ToDay investigated the efficacy of a tailored intervention using mobile technology to improve diet and PA behaviors, leading to weight loss in adults with a BMI ≥ 25 (Halse et al., 2019). Briefly, the larger study consisted of three randomized groups: (a) tailored feedback, (b) active control, and (c) online control. Only participants who were part of the tailored feedback arm wore both a Charge 2 and a GT3X+ concurrently, and therefore, only the data obtained from this group were included in the current analysis.

To be eligible for study enrollment, participants had to be between the ages of 18–65 years and classified as overweight or obese (BMI ≥ 25 to < 40 kg/m²), were required to have Internet access and own a smartphone (iPhone or Android), and had to complete the ToDay screening survey (Halse et al., 2019). Participants were also required to reside in metropolitan Perth, Western Australia, be able to attend the study center at Curtin University, and be able to engage in regular PA. It was essential that concurrent and valid data (minimum of 10 hr/day for 4 days) on both the Charge 2 and GT3X+ devices could be obtained. Based on these criteria, 75 participants enrolled in the tailored feedback arm of the larger study (Halse et al., 2019) were eligible for inclusion in the current study. Valid (aligned wear time) and complete (10 hr/day for 4+ days) data sets for the current study were available for 59 participants.

Procedure

The participants attended a 1-hr face-to-face visit, during which height (to the nearest centimeters) and body mass (to the nearest grams) were measured according to the standard protocol to calculate BMI (in kilograms per meter squared). During this visit, the participants were provided with a wrist-worn Fitbit Charge 2, together with a hip-worn ActiGraph GT3X+, and instructed on correct wear and use. This included concurrent wear of both devices continuously for 8 consecutive days, except when showering or engaging in aquatic activities. After the 8 days, the participants returned the ActiGraph GT3X+ to Curtin University via a prepaid envelope, and the data were subsequently downloaded. The participants continued to wear the Fitbit Charge 2 device over 12 months as part of the ToDay study, with device data continuously uploaded to an online, cloud-based platform (Fitabase, Fitabase LLC, San Diego, Ca) to facilitate researcher monitoring and analysis. Data relevant to the current analysis were extracted via an online, cloud-based platform (Fitabase).

ActiGraph GT3X+

The GT3X+ provides acceleration outputs in three planes of movement and a combined output of three dimensions (Straker & Campbell, 2012). The GT3X+ provides research-grade results in measuring energy expenditure, sedentary time, and varying intensity levels of PA (Brewer, Swanson, & Ortiz, 2017). The GT3X+ was programmed to record raw data at a frequency of 30 Hz, which were later reduced to vertical axis movement counts per 60-s epoch for the current analyses. The GT3X+ devices were initialized before the data collection, and then the data were downloaded

using ActiLife (version 6.0.0; ActiGraph, Pensacola, FL). A validated algorithm in SAS (version 9.3; SAS Institute, Cary, NC) was used to extract the waking wear time and time spent at each activity intensity, as well as the average accelerometer count at each activity intensity and raw step count per minute (nonfiltered, i.e., all steps counted; McVeigh et al., 2016). A single nonwear rule was applied to all of the data. All minutes in continuous periods of ≥ 90 min of 0 counts per minute (cpm), allowing for < 3 min with counts 1–50 cpm, were classed as nonwear. Common cut points (Matthews et al., 2008) were used to classify each minute as sedentary (< 100 cpm), light intensity (100–1,951 cpm), moderate intensity (1,952–5,724 cpm), or vigorous intensity ($> 5,724$ cpm) (Freedson, Melanson, & Sirard, 1998).

Fitbit Charge 2

The Charge 2 records PA through body motion using a microelectronic triaxial accelerometer (Feehan et al., 2018). This particular model can additionally track heart rate, using Fitbit PurePulse® technology (Fitbit Inc.,) (Haghighyegh, Khoshnevis, Smolensky, Diller, & Castriotta, 2019). All Charge 2 devices were initialized using the Fitbit online user interface, and data from the devices were extracted using the Fitabase software. Proprietary algorithms that translate activity count data into time spent in sedentary behavior, light PA, active minutes/MVPA, and step count (Reid et al., 2017) were used. Two csv files per participant were extracted from the Fitabase platform. These were the 60-s epoch “minute-IntensitiesNarrow” file and “minuteStepsNarrow” file. To be able to compare the Charge 2 data with the GT3X+ data, the same nonwear periods (as described for the GT3X+ data above) were applied to the Charge 2 data. The waking wear periods, which were extracted from the GT3X+ data (as described above), were used to identify the same periods for each participant on both of their Charge 2 files (intensities and steps). The waking wear periods were matched by choosing the “out of bed time” identified from the GT3X+ data as the start of the waking wear period and the “into bed time” identified from the GT3X+ data as the end of that waking wear period.

Data Reduction

Data from the GT3X+ and Charge 2 were manually matched for waking wear periods based on the GT3X+ data using Microsoft Excel (Microsoft, Redmond, WA). Thus, only valid wear time during waking hours that were simultaneously recorded on both devices were included for statistical analysis. Once the same periods were identified across the same days, the average value (per day) from each device, for step count, sedentary time, light PA, and MVPA were compared.

Statistical Analysis

To assess the convergent validity of the Charge 2, the device was compared with relevant GT3X+ outputs. These outputs included daily time spent in SB, light PA, MVPA, steps per day, and the number of “active” days (defined as days where ≥ 30 min of MVPA were taken). To assess the association between outputs, Pearson’s and Spearman’s correlation coefficients (for nonnormally distributed data), and intraclass correlation coefficients were used where appropriate. Correlations of $\geq .75$ implied excellent agreement, $.60 \pm .74$ implied good agreement, $.40 \pm .59$ implied fair agreement,

and $< .40$ implied poor agreement (Shrout & Fleiss, 1979). Paired t tests were used to assess for differences between the devices, and Bland–Altman plots were used to show the distribution of the differences. To determine relative differences, the mean absolute percentage error was used: $\text{MAPE} = \text{Abs} ([\text{ActiGraph measure} - \text{Fitbit measure}] / \text{ActiGraph measure}) \times 100$. Partial correlation coefficients were also used to assess the association between MAPE and age and BMI (after controlling for activity level). A univariate analysis of variance (controlling for activity level) was used to test for differences in the MAPE values between males and females. The proportion of agreement in the classification of “active” days was evaluated with a kappa coefficient. A κ value of $.21$ – $.40$ was considered “fair” agreement, $\kappa = .41$ – $.60$ was considered “moderate” agreement, $\kappa = .61$ – $.80$ was considered “substantial” agreement, and $\kappa \geq .81$ was considered “almost perfect agreement” (Landis & Koch, 1977). Statistical procedures were conducted using SPSS (version 22.0; SPSS Inc., Chicago, IL). The significance level accepted was $p < .05$.

Results

Table 1 shows the characteristics of the participants included in the current study. The participants ($n = 42$ females; $M \pm SD$: age = 48 ± 11 years) had a BMI of 31 ± 4 kg/m². Most participants were White (85%), and the majority had an annual income of between \$AUD100,000 and \$AUD149,000. The participants wore the devices for an average of 7.2 valid wear days and 15.6 hr/day waking wear time.

The Charge 2 outputs were significantly associated with the GT3X+ outputs. Time spent in SB, as measured by the two devices, had the weakest association and fair agreement, and step counts had the strongest association, with excellent agreement (Table 2).

For both the mean difference and relative difference, the Charge 2 recorded a significantly higher amount of time in SB, MVPA, and step count compared with the GT3X+, whereas the Charge 2 underestimated the time spent in light PA compared with the GT3X+ (Table 3, Figure 1). The median MAPE is reported in the table, as this is an outlier resistant measure.

The associations of age, gender, and BMI with the MAPE of each activity output were assessed (after controlling for relevant activity level). There was a significant positive correlation between BMI and MAPE of step count ($r = .277$), but there were no other significant associations for BMI or age. However, males had a significantly greater MAPE (12% higher for males, 95% confidence interval [2.72, 21.48]) for steps/day compared with females. Since no other significant associations were observed for gender, and these comparisons were not central to the intended analysis, the entire sample ($n = 59$) data were used for the remainder of the analysis.

Sedentary time had the greatest mean difference, but the smallest MAPE, contrasting MVPA, with the smallest mean difference but largest MAPE. The Bland–Altman plots (Figure 1) show the distribution of error for both estimates of sedentary time (Figure 1A), light PA (Figure 1B), MVPA (Figure 1C), and steps (Figure 1D). The wide limits of agreement suggest variation in differences for all outputs. The Bland–Altman plots showed that there was no apparent bias for the agreement and variances in sedentary time ($R^2 = .01$) or light PA ($R^2 = .02$) estimates between the two devices (Figure 1A and 1B). However, there was evidence for systematic bias for step count ($R^2 = .31$) and MVPA ($R^2 = .35$) between the two devices (Figure 1C and 1D). There was moderate agreement between the days classified as “active” on the two devices ($\kappa = .42$). A chi-squared test showed that there was

Table 1 Characteristics of Study Population

Variable	All (N = 59)	Males (n = 17)	Females (n = 42)
Age (years)	48 (11)	47 (10)	48 (11)
Height (cm)	169 (9)	179 (6)	164 (7)
Weight (kg)	89 (16)	100 (15)	85 (15)
Body mass index (kg/m ²)	31 (4)	31 (4)	31 (4)
Alcohol intake (n)			
Never	11	1	10
Monthly/less	10	1	9
2–4 times/month	11	3	8
2–3 times/week	17	8	9
4+ times/week	10	5	5
Ethnicity (n)			
White	50	16	34
Aboriginal	1	0	1
Asian	4	0	4
Pacific Islander	0	0	0
Black	0	0	0
Mixed	4	2	2
Household income, \$AUD (n)			
<\$29,999	3	1	2
\$30,000–\$59,000	4	0	4
\$60,000–\$99,999	12	2	10
\$100,000–\$149,999	16	4	12
\$150,000–\$199,999	15	7	8
\$200,000 or more	8	4	4
Don't know	0	0	0
Prefer not to answer	1	0	1

Note. Data are presented as mean (SD) or n.

Table 2 Association of Activity Behavior Measured by the Fitbit Charge 2 and ActiGraph GT3X+ (N = 59)

Variable	Pearson/ Spearman correlation (r)	Intraclass correlation (r [95 % CI])
Sedentary time (min/day)	.507	.504 [.287, .672]
Light PA time (min/day)	.788	.785 [.663, .866]
MVPA time ^a (min/day)	.640	.641 [.462, .770]
Steps/day (n/day)	.946	.925 [.877, .955]

Note. CI = confidence interval; MVPA = moderate to vigorous physical activity; PA = physical activity.

^aSpearman rank correlation used for MVPA due to skewed nature of data.

a significant difference in the proportion of participants classified as participating in an average of ≥ 30 min of MVPA per day ($\chi^2 = 10.80$, $p < .001$) between Charge 2 ($n = 38$, 64%) and GT3X+ ($n = 31$; 52%).

Discussion

This study presents novel data that demonstrate the convergent validity of the Fitbit Charge 2 with the ActiGraph GT3X+ in

Australian adults with a BMI ≥ 25 . While there were substantial associations between both devices for all measures (i.e., SB, light PA, MVPA, and step count), there were also mean and relative (MAPE) differences, which may have implications that need to be considered more carefully for populations who are already at risk of being highly sedentary and insufficiently active (e.g., overweight/obese adults).

There was a moderate correlation and low relative error between the two devices for sedentary time, with the Charge 2 overestimating sedentary time by approximately half an hour per day compared with the GT3X+. This result is of particular relevance to very sedentary populations, where the accurate estimation of time spent in low-intensity behaviors is important. The findings in the current study are similar to those of Redenius et al. (2019), who used similar cut points for their GT3X+ processing as in the current study, and reported that in participants of a healthy weight range, the wrist-worn Fitbit Flex overestimated time in sedentary behavior compared to the waist-worn GT3X+ with a mean difference of 37 min/day. In a more sedentary population (but with participants of similar BMIs to the participants in the current study), the Charge 2 overestimated sedentary time by more than 2 hr/day in older adults with knee osteoarthritis (Collins et al., 2019). Mikkelsen et al. (2020) reported a lower mean measure of SB by the Charge 2 (−25 min/day) compared with the GT3X+. Possible reasons for the discord in these findings include

Table 3 Absolute and Relative Differences of Activity Behavior Measured by the Fitbit Charge 2 and ActiGraph GT3X+ (*n* = 59)

Variable	Fitbit Charge 2 Mean (SD)	ActiGraph GT3X+ Mean (SD)	Mean difference [95% CI]	Mean percent error [95% CI]	Relative difference MdAPE% [95% CI] ^a
Sedentary time (min/day)	635 (90)	598 (81)	37 [15, 60]	7.2 [3.2, 11.3]	6.8 [1.8, 19.4]
Light PA time (min/day)	261 (67)	294 (73)	-32 [-20, -44]	-10.0 [-13.98, -6.08]	13.1 [8.8, 14.1]
MVPA time (min/day)	54 (46)	39 (29)	15 [6, 23]	60.4 [29.7, 91.1]	26.5 [6.7, 45.0]
Steps/day (<i>n</i> /day)	10,068 (4,184)	8,283 (3,372)	1,812 [1,430, 2,195]	23.0 [18.5, 27.6]	23.0 [16.2, 29.3]

Note. MAPE = mean absolute percentage error; MdAPE = median MAPE; CI = confidence interval; MVPA = moderate to vigorous physical activity; PA = physical activity.

^aThe MdAPE is reported, as this is an outlier-resistant measure.

Figure 1 — Bland–Altman plot of differences in (a) sedentary time, (b) light PA time, (c) MVPA time, and (d) steps between the Fitbit Charge 2 and ActiGraph GT3X+ against the mean of Charge 2 and GT3X+. The solid line represents the mean of the differences between devices, and the dotted lines are 95% limits of agreement (± 1.96 SD). Note: Two individuals had very high daily activity and with these removed the regression was still significant (for steps and MVPA), though a little weaker. MVPA = moderate to vigorous physical activity; PA = physical activity.

differences in the processing by Mikkelsen et al. (2020), such as different cut points, use of vector magnitude, and use of a self-report log for aligning the wear time between devices.

In the current study, the Charge 2 underestimated the time spent in light-intensity PA by approximately 30 min/day when compared with the GT3X+. This finding is in contrast to a study where the wrist-worn Fitbit Flex provided similar estimates of time spent in light PA in comparison with the GT3X+ under community-dwelling conditions (Reid et al., 2017). In the current study, the Fitbit algorithm appears to have classified some light PA time as

sedentary time. This difference could be attributed to the added heart rate monitoring associated with the Charge 2 device, which may have improved the classification of time spent in varying levels of intensity (Stahl et al., 2016).

Results of the current study show that the Charge 2 recorded a significantly higher amount of time in MVPA than the GT3X+ (15 min/day). Studies with other models of the Fitbit (i.e., Charge and Fitbit Flex) have also reported an overestimation of MVPA (up to 1 hr more per day) in comparison with the GT3X+ (Feehan et al., 2018; Redenius et al., 2019; Reid et al., 2017). Reid et al. (2017)

reported that both models of Fitbits (One and Flex) included in their study overestimated the time spent in all intensities of PA and showed that, as PA intensity increases, Fitbit estimates become less accurate. As a proportion of the waking wear day, the 15-min overestimation observed in the current study was associated with the largest MAPE and widest limits of agreements. Additionally, there is some evidence of systematic bias in the current study, suggesting that the discrepancy tends to increase as the total mean daily MVPA volume increases. In contrast to the current study, Collins et al. (2019) reported that time spent in MVPA was underestimated by the Fitbit Charge 2 compared with the GT3X+ by 5 min/day. Importantly, their participants had a slower walking speed (79 steps/min) than the established 100 steps/min guideline (Collins et al., 2019). Therefore, their results may be attributable to limited engagement in MVPA, resulting from slower walking speeds, physical function limitations, and knee pain associated with knee osteoarthritis (Collins et al., 2019). In overweight participants, Mikkelsen et al. (2020) also reported lower estimates of MVPA on the Charge 2 (59 min/day) compared with the GT3X+ (90 min/day), but the very high estimate of daily MVPA for the participants (three times the daily MVPA guideline) in Mikkelsen's study may be explained either by that study having recruited an exceptionally physically active population or by GT3X+ data processing issues accounting for this discrepancy.

The findings of this study highlight the implications of using consumer activity trackers in pragmatic/community settings with participants who may be overweight/obese. In the current study, age did not appear to have an important influence on the differences in any activity measures between the Charge 2 and GT3X+. However, the MAPE in the step count (between the two devices) tended to increase as BMI increased. This may be due to the step count algorithm in the GT3X+ being affected by abdominal obesity (increasing tilt on the hip-worn device and affecting vertical axis output; Belanger et al., 2014; Feito et al., 2012) or due to an effect of obesity on the wrist-worn Fitbit device. This may be an important factor for future studies to consider. The difference in step count measurement between the Charge 2 and GT3X+ was larger for males than for females. Though the male sample size was only 17, BMI and sex-related differences for step count measures between devices may need to be considered in future studies.

The results of the current study confirm previous findings that, despite the strong correlation between devices for step count, the Charge 2 overestimates step count compared with research-grade accelerometers (Chu et al., 2017; Collins et al., 2019; Mikkelsen et al., 2020) and that the overestimation may be worse in people with high BMI values (Mikkelsen et al., 2020). In the current study, the Charge 2 overestimated the step count by an average of 1,812 steps per day, and there was evidence of this worsening as the total step count increased. This overestimation of the daily step count resulted in approximately 12% more of the participants being classified as "active" (i.e., achieving ≥ 30 min MVPA per day) compared with the GT3X+. For interventions, the usefulness of wearable activity trackers is to be able to provide real-time feedback to participants on their PA behavior. Therefore, improving the estimations from these devices is important for providing accurate feedback to participants. Also, the public health implications of consumers who are using these devices being identified as sufficiently active when, in fact, they are not is an important consideration. Using specific parameters of the population being studied may be important to consider if wearable activity trackers are being used to support behavior change in an overweight/obese population.

There are several strengths to our study. To our knowledge, this is one of the first studies to report the convergent validity of the Fitbit Charge 2 in an overweight/obese adult population. Second, the study was completed under community-dwelling conditions to more accurately represent unstructured lifestyle activities. Third, the study examined different components of PA with varying levels of intensity with a Fitbit device that monitors heart rate, which can aid in producing a more accurate recording of PA intensities (O'Driscoll et al., 2020; Stahl et al., 2016). The use of the GT3X+ as the comparative accelerometer device allows for our findings to be relevant to other studies in which the GT3X+ has been used to provide comparisons between other wearable activity trackers. However, the cut points used for classifying thresholds of activity from the GT3X+ data (in the current study) were developed in a young and healthy population (Sasaki, John, & Freedson, 2011) and may not be as accurate for a free-living population with a BMI ≥ 25 . Additionally, the GT3X+ is limited in differentiating sitting versus standing (An, Kim, & Lee, 2017), and this may have influenced the current study's findings. Other limitations include the different attachment sites of the devices, with the Charge 2 being a wrist-worn device and the GT3X+ being a waist-worn device. With the added movement of the arm/forearm in other day-to-day activities, the Charge 2 may have mistakenly recorded those movements as steps or PA engagement and could be a factor resulting in overestimations in both MVPA and step count (Chu et al., 2017). Although this study did not assess the effect of tilt (due to midsection obesity) on the GT3X+ output, there is evidence from the literature to show that BMI and tilt angle may have an effect on the step count accuracy and count values recorded by the ActiGraph (Belanger et al., 2014). Finally, the Charge 2 device is also no longer commercially available and has now been replaced by the Charge 3. The unknown proprietary hardware and software changes to subsequent models may result in different levels of accuracy than what has been tested in this study.

Conclusion

The addition of heart rate monitoring associated with the Fitbit Charge 2, together with the ability to monitor and download participant data remotely and in real time, and the ease of use in a community setting suggest that the Fitbit Charge 2 may be a viable and pragmatic tool to support behavior change in populations with a BMI ≥ 25 . The Fitbit Charge 2 demonstrated very good agreement with the research-grade ActiGraph GT3X+ for low-intensity activities and fair to moderate agreement with higher intensity activities. However, caution is advised for use in research studies where absolute measurement estimates are important, given the potential error in activity estimates for adults with a BMI ≥ 25 .

Acknowledgments

The authors sincerely thank all participants from the ToDay study. Funding for the ToDay study was provided by a Healthway Health Promotion Research Grant and the East Metropolitan Health Service. Funding for Fitabase was provided by a Curtin Institute of Computation grant. The sponsors had no role in the design of the study; collection, analyses, or interpretation of data; writing of the manuscript; and decision to publish the results.

References

- An, H.-S., Kim, Y., & Lee, J.-M. (2017). Accuracy of inclinometer functions of the activPAL and ActiGraph GT3X+: A focus on

- physical activity. *Gait and Posture*, 51, 174–180. PubMed ID: 27780084 doi:10.1016/j.gaitpost.2016.10.014
- Belanger, M.L., Kestens, Y., Gilbert, J.A., Tremblay, A., & Mathieu, M.E. (2014). Interaction between body weight status and walking speed in steps monitoring by GT3X accelerometer. *Applied Physiology, Nutrition, and Metabolism*, 39(8), 976–979. PubMed ID: 24823315
- Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D.C., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One*, 13(2), e0192691. PubMed ID: 29489850 doi:10.1371/journal.pone.0192691
- Brewer, W., Swanson, B.T., & Ortiz, A. (2017). Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. *BMJ Open Sport & Exercise Medicine*, 3(1), e000254. PubMed ID: 29018543 doi:10.1136/bmjsem-2017-000254
- Cadmus-Bertram, L.A., Marcus, B.H., Patterson, R.E., Parker, B.A., & Morey, B.L. (2015). Randomized trial of a Fitbit-based physical activity intervention for women. *American Journal of Preventive Medicine*, 49(3), 414–418. PubMed ID: 26071863 doi:10.1016/j.amepre.2015.01.020
- Chu, A.H., Ng, S.H., Paknezhad, M., Gauterin, A., Koh, D., Brown, M.S., & Muller-Riemenschneider, F. (2017). Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. *PLoS One*, 12(2), e0172535. PubMed ID: 28234953 doi:10.1371/journal.pone.0172535
- Collins, J.E., Yang, H.Y., Trentadue, T.P., Gong, Y., & Losina, E. (2019). Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. *PLoS One*, 14(1). doi:10.1371/journal.pone.0211231
- Evenson, K.R., Goto, M.M., & Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 159. doi:10.1186/s12966-015-0314-1
- Feehan, L.M., Geldman, J., Sayre, E.C., Park, C., Ezzat, A.M., Yoo, J.Y., ... Li, L.C. (2018). Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth*, 6(8), e10527. PubMed ID: 30093371 doi:10.2196/10527
- Feito, Y., Bassett, D.R., Thompson, D.L., & Tjo, B.M. (2012). Effects of body mass index on step count accuracy of physical activity monitors. *Journal of Physical Activity and Health*, 9(4), 594–600. PubMed ID: 21946229 doi:10.1123/jpah.9.4.594
- Feito, Y., Bassett, D.R., Tjo, B., & Thompson, D.L. (2011). Effects of body mass index and tilt angle on output of two wearable activity monitors. *Medicine & Science in Sports & Exercise*, 43(5), 861–866. PubMed ID: 20962689 doi:10.1249/MSS.0b013e3181fef4d0
- Freedson, P.S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, 30(777), 781.
- Haghighyegh, S., Khoshnevis, S., Smolensky, M.H., Diller, K.R., & Castriotta, R.J. (2019). Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 21(11), e16273. PubMed ID: 31778122 doi:10.2196/16273
- Halse, R.E., Shoneye, C.L., Pollard, C.M., Jancey, J., Scott, J.A., Pratt, I.S., ... Kerr, D.A. (2019). Improving nutrition and activity behaviors using digital technology and tailored feedback: Protocol for the LiveLighter Tailored Diet and Activity (ToDAY) randomized controlled trial. *JMIR Research Protocols*, 8(2):e12782. doi:10.2196/12782
- Hargens, T.A., Deyarmin, K.N., Snyder, K.M., Mihalik, A.G., & Sharpe, L.E. (2017). Comparison of wrist-worn and hip-worn activity monitors under free living conditions. *Journal of Medical Engineering and Technology*, 41(3), 200–207. PubMed ID: 28078908 doi:10.1080/03091902.2016.1271046
- Hayes, A., Lung, T., Bauman, A., & Howard, K. (2017). Modelling obesity trends in Australia: Unravelling the past and predicting the future. *International Journal of Obesity*, 41(1), 178–185. PubMed ID: 27671035 doi:10.1038/ijo.2016.165
- Hickey, A.M., & Freedson, P.S. (2016). Utility of consumer physical activity trackers as an intervention tool in cardiovascular disease prevention and treatment. *Progress in Cardiovascular Diseases*, 58(6), 613–619. PubMed ID: 26943981 doi:10.1016/j.pcad.2016.02.006
- Huang, Y., Xu, J., Yu, B., & Shull, P.B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait and Posture*, 48, 36–41. PubMed ID: 27477705 doi:10.1016/j.gaitpost.2016.04.025
- Júdice, P.B., Santos, D.A., Hamilton, M.T., Sardinha, L.B., & Silva, A.M. (2015). Validity of GT3X and Actiheart to estimate sedentary time and breaks using ActivPAL as the reference in free-living conditions. *Gait and Posture*, 41(4), 917–922. PubMed ID: 25852024 doi:10.1016/j.gaitpost.2015.03.326
- Landis, J.R., & Koch, G.G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, (2), 363–374. doi:10.2307/2529786
- Matthews, C.E., Chen, K.Y., Freedson, P.S., Buchowski, M.S., Beech, B.M., Pate, R.R., & Troiano, R.P. (2008). Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American Journal of Epidemiology*, 167(7), 875–881. PubMed ID: 18303006 doi:10.1093/aje/kwm390
- McVeigh, J.A., Winkler, E.A., Healy, G.N., Slater, J., Eastwood, P.R., & Straker, L.M. (2016). Validity of an automated algorithm to identify waking and in-bed wear time in hip-worn accelerometer data collected with a 24 h wear protocol in young adults. *Physiological Measurement*, 37(10), 1636–1652. PubMed ID: 27652717 doi:10.1088/0967-3334/37/10/1636
- Mikkelsen, M.K., Berg-Beckhoff, G., Frederiksen, P., Horgan, G., O'Driscoll, R., Palmeira, A.L., ... Larsen, S.C. (2020). Estimating physical activity and sedentary behaviour in a free-living environment: A comparative study between Fitbit Charge 2 and Actigraph GT3X. *PLoS One*, 15(6), e0234426. PubMed ID: 32525912 doi:10.1371/journal.pone.0234426
- O'Driscoll, R., Turicchi, J., Beaulieu, K., Scott, S., Matu, J., Deighton, K., ... Stubbs, J. (2020). How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine*, 54(6), 332–340. PubMed ID: 30194221
- Redenius, N., Kim, Y., & Byun, W. (2019). Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. *BMC Medical Research Methodology*, 19(1), 29. PubMed ID: 30732582 doi:10.1186/s12874-019-0668-1
- Reid, R.E.R., Insogna, J.A., Carver, T.E., Comptour, A.M., Bewski, N.A., Sciortino, C., & Andersen, R.E. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of Science and Medicine in Sport*, 20(6), 578–582. PubMed ID: 27887786 doi:10.1016/j.jsams.2016.10.015
- Roberto, C.A., Swinburn, B., Hawkes, C., Huang, T.T., Costa, S.A., Ashe, M., ... Brownell, K.D. (2015). Patchy progress on obesity prevention: Emerging examples, entrenched barriers, and new thinking. *The Lancet*, 385(9985), 2400–2409. doi:10.1016/S0140-6736(14)61744-X
- Sasaki, J.E., John, D., & Freedson, P.S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in*

- Sport*, 14(5), 411–416. PubMed ID: [21616714](#) doi:[10.1016/j.jsams.2011.04.003](#)
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. PubMed ID: [18839484](#) doi:[10.1037/0033-2909.86.2.420](#)
- Stahl, S.E., An, H.-S., Dinkel, D.M., Noble, J.M., & Lee, J.-M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & Exercise Medicine*, 2(1), e000106. PubMed ID: [27900173](#) doi:[10.1136/bmjsem-2015-000106](#)
- Straker, L., & Campbell, A. (2012). Translation equations to compare ActiGraph GT3X and Actical accelerometers activity counts. *BMC Medical Research Methodology*, 12(1), 54. doi:[10.1186/1471-2288-12-54](#)
- Sushames, A., Edwards, A., Thompson, F., McDermott, R., & Gebel, K. (2016). Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One*, 11(9). doi:[10.1371/journal.pone.0161224](#)
- Toth, L.P., Park, S., Pittman, W.L., Sarisaltik, D., Hibbing, P.R., Morton, A.L., . . . Bassett, D.R. (2018). Validity of activity tracker step counts during walking, running, and activities of daily living. *Translational Journal of the American College of Sports Medicine*, 3(7), 52–59.
- Tully, M.A., McBride, C., Heron, L., & Hunter, R.F. (2014). The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Research Notes*, 7(1), 952. doi:[10.1186/1756-0500-7-952](#)
- Yoost, J., Gerlach, J., Sinning, M., & Cyphert, H. (2018). The use of fitbit technology among rural obese adolescents. *Journal of Obesity and Nutritional Disorders: JOND-122*. doi:[10.29011/JOND-122.100022](#)
- Yusefzadeh, H., Rahimi, B., & Rashidi, A. (2019). Economic burden of obesity: A systematic review. *Journal of Health and Social Behavior*, 2(1), 7–12. doi:[10.4103/SHB.SHB_37_18](#)
- Zenko, Z., Willis, E.A., & White, D.A. (2019). Proportion of adults meeting the 2018 physical activity guidelines for Americans according to accelerometers. *Frontiers in Public Health*, 7, 135. PubMed ID: [31231627](#) doi:[10.3389/fpubh.2019.00135](#)