Sensitivity of the Misfit Shine™ to Detect Changes in Laboratory-Based and Free-Living Physical Activity

Greg Petrucci Jr., Patty Freedson, Brittany Masteller, Melanna Cox, John Staudenmayer, and John Sirard
University of Massachusetts

Purpose: Determine the sensitivity of the Misfit Shine™ (MS) to detect changes in physical activity (PA) measures (steps, “points,” kCals) in laboratory (LAB) and free-living (FL) conditions. **Methods:** Twenty-one participants wore the MS and ActiGraph GT3X+™ accelerometer (AG) at the hip and dominant-wrist during three, one-hour LAB sessions: sedentary (SS), sedentary plus walking (SW), and sedentary plus jogging (SJ). Direct observation (DO) of steps served as the criterion measure. Devices were also worn during two FL conditions: 1) active week (ACT) and 2) inactive week (INACT). For LAB and FL, significant differences were examined using paired t-tests and linear mixed effects models, respectively. Linear mixed effects models were used to estimate differences between MS estimated steps and DO ($\alpha \leq 0.05$). **Results:** For all hip-worn MS measures and wrist-worn MS estimates of steps and “points,” there was a significant increase ($p < .05$) from SS to SJ. However, wrist-worn MS kCal estimates were greater for SJ, compared to SS and SW, which were similar to each other (95% CI [95.5, 152.8] and [141.1, 378.9], respectively). Compared with DO, MS hip significantly underestimated steps by 3.5%, while MS wrist significantly overestimated steps by 4.2%. During FL conditions, all MS measures were sensitive to changes between ACT and INACT ($p < .0001$). **Conclusion:** Although there were systematic errors in step estimates from the MS, it was sensitive to changes during LAB and FL, and may be a useful tool for interventionists where tracking changes in PA is an important exposure or outcome variable.

Keywords: consumer activity tracker, steps, validation

Low prevalence of physical activity (PA) and high levels of sedentary behavior (SB) are independent public health concerns (Greer, Sui, Maslow, Greer, & Blair, 2015; Maher, Mire, Harrington, Staiano, & Katzmarzyk, 2013). In response to this problem, considerable research efforts have focused on validating research-grade accelerometers to objectively quantify the relationships among PA, SB, and health (Matthews et al., 2008; Matthews et al., 2016; Troiano, McClain, Brychta, & Chen, 2014). Recently, similar technology and analytics initially used in research-grade accelerometers have been integrated into consumer activity trackers.

Noted as the top fitness movement in the American College of Sports Medicine Annual Survey of Fitness Trends, the market for wearable fitness trackers market is currently valued at approximately \$21 billion and is expected to increase to \$90 billion by the end of 2028 (Future Market Insights, 2017; Thompson, 2015). Researchers have shown considerable interest in studying the validity, reliability, and public health benefits of consumer activity trackers (Evenson, Goto, & Furberg, 2015; Wright, Hall Brown, Collier, & Sandberg, 2017). An increasing number of consumer activity trackers now offer multiple wear location options, prolonged battery life, and Bluetooth capabilities all of which can be attractive features for both researchers and the consumer (Hickey & Freedson, 2016). The Misfit Shine™ (MS; Fossil Group, Richardson, TX, USA) is one of Fossil’s activity tracker devices that offers these attributes. In addition, it is one of the few devices that is water proof and has a battery life of six months. Both of

these additional features are potentially attractive to researchers as methods to decrease missing data by improving wear compliance and not relying on participants to put the device back on after swimming or bathing. Fossil’s activity trackers have been well received by the public as the MS was recently listed as a top-rated activity tracker (Stables, 2018).

A recent laboratory study showed estimates of steps/minute from a hip-worn MS were similar to step estimates from a hip-worn ActiGraph GT3X+™ (AG). In addition to demonstrating strong concurrent validity, the MS steps/minute estimates were similar across three different wear locations (hip, wrist, foot) (Mendoza, Hickey, & Freedson, 2015). Intraclass correlations between repeated treadmill conditions were 0.91, 0.92, and 0.98 for hip, wrist, and ankle wear-locations, respectively (Hickey, Mendoza, & Freedson, 2015). The ability of the MS to produce increased step counts with increased intensity suggests it may be sensitive in detecting change in activity over time. Another Misfit model has shown reliability in estimating steps during repeated treadmill walking bouts at 3.2 km·h⁻¹, 4.82 km·h⁻¹, and 6.42 km·h⁻¹ (Fokkema, Kooiman, Krijnen, Van Der Schans, & Groot, 2017). However, before widespread adoption of the MS in research, as an intervention tool or as a replacement for research-grade measurement devices, it is important to directly determine if the device can detect change in measures of activity behavior.

Previous work has studied the sensitivity of research-grade accelerometers to detect change (Lee, Clark, Winkler, Eakin, & Reeves, 2015; Montoye, Pfeiffer, Sutton, & Trost, 2014; Swartz, Rote, Cho, Welch, & Strath, 2014), but this remains unknown for consumer activity trackers. Using the AG to establish convergent validity, the primary aim of this study was to extend previous validation work using the MS (Fokkema et al., 2017; Mendoza

Petrucci, Masteller, Cox, and Sirard are with the Dept. of Kinesiology; Freedson is Professor Emerita with the with the Dept. of Kinesiology; Staudenmayer is with the Dept. of Mathematics and Statistics; University of Massachusetts Amherst, Amherst, MA. Petrucci (gpetrucci@umass.edu) is corresponding author.

et al., 2015) by examining the sensitivity of the MS to detect changes during 1) a structured laboratory-based protocol to simulate sedentary office work interrupted by bouts of activity, and 2) a two-week free-living protocol, where participants were inactive for one week and for the second week, they were instructed to achieve the 2008 Physical Activity Guidelines (Office of Disease Prevention and Health Promotion, 2008). A secondary aim was to assess criterion validity of MS step estimates during the lab condition, using direct observation (DO) with manual step counting as the criterion step measure.

Methods

Participants

Participants were recruited locally through word of mouth, flyers, and email distributions. Participants were between the ages of 18 and 40 years, in good health, and participated in regular PA. Prior to data collection, a trained researcher screened potential participants for any physical limitations that could affect their ability to engage in PA, and answered questions pertaining to study protocols. This study was approved by the University of Massachusetts Institutional Review Board. If deemed eligible, participants read and signed an informed consent document. It should be noted that not all participants completed both laboratory-based and free-living conditions. Participant characteristics for both samples are presented in Table 1.

Design

This study consisted of two distinct conditions: 1) laboratory-based, simulated free-living condition (LAB) and 2) free-living condition (FL).

Instruments and Measures

After consenting, height and weight, to the nearest 0.25 cm and 0.1 kg, were measured by trained research staff using a standard floor stadiometer (Weigh and Measure, LLC; Onley, MD, USA) and physicians' scale (Detetco; Webb City, MO, USA).

Direct Observation. A direct observation (DO) methodology for step counting was used as the criterion measure of steps during the LAB condition. The DO step counting was not completed for the sedentary session (SS) since all of the participants remained seated for the entire session. The participant's feet were videotaped during the treadmill walking and jogging portions of the sedentary plus walking session (SW) and sedentary plus jogging (SJ) LAB sessions using a GoProHERO+LCD™ camera (GoPro, Inc., San Mateo, CA, USA). Subsequently, two independent researchers viewed the videos and counted each step using the Observer XT software (Noldus Information Technology; Wageningen, Netherlands). The Observer XT software permitted step counters to count

steps in slow motion, rewind, and use a keyboard shortcut key to identify each step. Steps in all videos were counted twice and discrepancies in step counts that differed by greater than 1% were counted a third time. The average between two DO step counts that differed by less than 1% was used for subsequent analysis.

ActiGraph Accelerometer. The ActiGraph GT3X+™ accelerometer (ActiGraph, LLC, Pensacola, FL; AG) was used to assess concurrent validity of MS PA outcomes. The AG is a small ($4.6 \times 3.3 \times 1.5$ cm), lightweight (19 g), triaxial accelerometer used in many research applications (Matthews et al., 2008; Troiano et al., 2008). The AG step counting function has shown high accuracy for measuring steps in the laboratory (compared with DO ICC: 0.72–0.99) and in free-living settings (compared with Yamax Digiwalker ICC: 0.90) (Lee, Williams, Brown, & Laurson, 2015).

The AG includes a micro-electro-mechanical system (MEMS) based accelerometer which captures data within a dynamic range of ± 6 G-forces. The acceleration data are sampled by a 12-bit analog to digital converter at rates ranging from 30 Hz to 100 Hz and stored in a non-filtered accumulated format (G-forces). These data are stored directly into non-volatile flash memory. For the present study, raw data were collected at 80 Hz and post-processed in the ActiLife software (Version 6.13.3). The ActiLife software allows users to generate files containing any desired combination of parametric data (e.g., 1-s epoch, 60-s epoch) during the data management step (ActiGraph, 2012). The wrist AG was positioned midway between the radial and ulnar styloid processes using a Velcro strap. The hip AG was positioned on the midline of body, centered above the right knee using the manufacturer belt.

Misfit Shine. The Misfit Shine (MS) is a consumer activity tracker, released in August 2013, which estimates steps, points, and Calories (kCals) using proprietary algorithms. The MS was worn on the dominant wrist and hip. The dominant wrist site was chosen because that is where the manufacturer recommends the MS to be worn. We chose to also put the AG on the dominant wrist to ensure that differences in sensitivity between MS and AG devices were not a function of different wear locations. The wrist MS was positioned distal to the elbow, relative to the wrist AG, using the manufacturer wrist strap. The hip MS was positioned medially to the hip AG on the belt using a magnetic clip provided by the manufacturer. To set up the device, a user inputs height, weight, sex, and date of birth. A research assistant input participant-specific information before the start of each condition. Users can sync the MS to a Bluetooth capable device at any time to see their daily progress.

Conditions

Laboratory Condition. The LAB condition consisted of three, one-hour sessions: sedentary session (SS), sedentary plus walking session (SW), and a sedentary plus jogging session (SJ). For all sessions, participants wore MS and AG devices on the right hip and on the dominant wrist. Sessions were presented in a balanced order to diminish the effects of order on participant behavior and device readings.

For the SS session, participants were asked to sit quietly for 60 minutes, performing sedentary activities such as sitting, computer work, or using their cell phone. The purpose of this session was to simulate a state of minimal activity or movement (sedentary office work). All participants remained seated for the entire SS session.

The SW and SJ sessions consisted of two components. First, participants engaged in 30 minutes of sedentary time (same as SS). Second, participants completed 30 minutes of treadmill walking and

Table 1 Participant Demographics by Condition

Condition	N	Gender (% Female)	Age (Years)	BMI ($\text{kg}\cdot\text{m}^{-2}$)
Laboratory	21	46	24.8 $\pm$ 3.57	24.2 $\pm$ 3.45
Free-living	21	48	24.5 $\pm$ 3.74	23.6 $\pm$ 3.77

Note. 68% of participants completed both conditions. Age and BMI values listed are $M \pm SD$.

jogging at $5.14 \text{ km}\cdot\text{hr}^{-1}$ and $8.04 \text{ km}\cdot\text{hr}^{-1}$ for SW and SJ, respectively. MS and AG PA outcomes for the SW and SJ sessions were the sum of the first and second 30-minute components. DO step counting was conducted for the SW and SJ sessions and served as the criterion measure of steps.

Free-Living Condition. The FL condition of the study took place over two seven-day periods: 1) an active week (ACT) and 2) an inactive week (INACT). Participants wore MS and AG devices at the hip and dominant wrist for both weeks. Participants were instructed to wear devices during waking hours and only take them off if doing activity where they would get wet. Participants were also instructed to put on and remove all devices at the same time. The AG was initialized to start collecting data at 4:00 AM on the morning of the first day of the FL condition (day after participants received devices from researchers) and data were collected until the devices were returned after ACT and INACT were completed. The ACT and INACT conditions were balanced across subjects to account for possible order effects on participant behavior and device readings.

For the ACT week, researchers instructed participants to engage in 150 minutes of moderate, or 75 minutes of vigorous intensity, locomotor PA (e.g., walking or running) to meet the 2008 PA Guidelines ([Office of Disease Prevention and Health Promotion, 2008](#)). Participants were all recreationally active and were familiar with minutes of moderate-to-vigorous PA (MVPA). To ensure clarification of these guidelines, research assistants provided examples for meeting these guidelines to participants. During the INACT week, participants were instructed to refrain from any purposeful activity.

Data Management

Data management, processing, and analysis were performed using ActiLife software (Version 6.13.3) and R, an open-source computing language and statistics package ([R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2017](#)).

Laboratory Condition. For the LAB condition, raw AG data collected at 80 Hz were post-processed and aggregated in 1-second epochs using ActiLife software (Version 6.13.3) to obtain steps and vertical counts per second. These second-by-second data were imported into R and a custom program was written to calculate the following AG PA outcomes for SS, SW, and SJ: total steps, kCals and counts per session. Total steps and counts (vertical axis) per session were computed as the sum of second-by-second data derived from the respective proprietary ActiLife algorithm. For the AG hip and wrist, total kCals per session were derived using the Freedson 1998 equation which was used to first calculate METs ([Freedson, Melanson, & Sirard, 1998](#)). METs were subsequently multiplied by participant body weight and duration to compute total kCals per session.

Specific syncing procedures were used to record MS LAB data. Prior to the start of the SS session and each 30-minute component of SW and SJ sessions, participants were asked to remain still for three minutes. A three-minute wait period was chosen because it proved to be an adequate amount of time for the devices to communicate data to the MS app during our pilot sessions. After three minutes, “pre-activity” data from the MS were recorded, and then the session/component began. Similarly, at the end of each session/component participants were asked to remain still for three minutes before researchers synced the devices,

to obtain “post-activity” data. The difference between “post-activity” and “pre-activity” measures defined the activity metrics for that session/component. This method was used to determine total MS steps, kCals, and points per session for the LAB condition. Points is a unit-less summary measure of overall movement that is calculated using proprietary algorithms. We cannot say for certain what this measure represents, however it is a primary outcome that MS reports and we hypothesized it to be positively associated with increased PA. These data were recorded in a spreadsheet and imported into R for statistical evaluation.

Free-Living Condition. For the FL condition, raw AG data collected at 80 Hz were post-processed and aggregated in 1-minute epochs using ActiLife software (Version 6.13.3) to obtain steps and vertical-axis counts per minute. These data were imported into R and a custom program was written to calculate the following AG PA outcomes for ACT and INACT: 1) total steps/day, 2) total kCals/day, and 3) total counts/day. Similar methods to those used for the LAB condition were employed to derive FL AG outcomes; however, summary data were derived from 1-minute epoch data. We chose to look at 24 hours of AG data (e.g., no wear time filter) because the MS does not report time-stamped data and there would be no way to filter non-wear data for this device. We analyzed minutes spent in MVPA during ACT and INACT. Minutes of MVPA from the hip-worn AG were defined by using the Freedson cutpoint ([Freedson et al., 1998](#)).

For the FL MS data, devices were synced to the MS mobile app upon completion of the ACT and INACT weeks. Total steps/day, kCals/day, and points/day were recorded from the MS mobile app, entered into a spreadsheet, and imported to R for statistical evaluation.

Statistical Evaluation

All statistical evaluation was done using R software programs. To determine the sensitivity of MS and AG to detect change in activity among LAB conditions, paired t-tests were used for unique combinations of PA outcome and wear-location, for all sessions. To determine the sensitivity of MS and AG to detect change between FL conditions (ACT week versus INACT week), repeated measures linear mixed effects models for each PA outcome and wear-location combination were calculated. An alpha level of 0.05 was used to define significant differences for all comparisons.

To assess criterion validity of MS step estimates during the SW and SJ LAB sessions, linear mixed effects models were fit for the difference (bias) between device estimated steps (both sessions combined) and DO step counts. Both SW and SJ sessions were combined together because the degree of overestimation and underestimation was similar between sessions. Separate models were fit for the differences between MS device estimates and DO step counts at each wear location. Mean bias and 95% bias intervals that did not span zero defined significant differences between MS estimates and DO. Percent difference between MS and DO steps were also examined to quantify the magnitude and direction of the bias. Direct comparisons between MS and AG steps were not made because the AG is not a criterion measure for steps ([Bassett, Rowlands, & Trost, 2012](#)).

Results

During the LAB and FL conditions there were several occasions when the AG and/or the MS did not record data. During 32%

(34/105) of the LAB conditions, the MS kCals estimate displayed on the app was lower at the completion of the condition, compared with the start. These data points were removed from further analyses. We believe this error is related to prolonged time for the device to sync with the app. We had sufficient pilot data to suggest that a three-minute wait period would be enough time for devices to sync, however the device did not consistently update during the LAB conditions in our study. These errors were not present during the FL condition. On one occasion during the FL condition the AG did not record data.

Figure 1 shows means (95% confidence intervals [CI]) for hip- and wrist-worn MS and AG PA outcomes for each LAB session. In addition, the DO step count means (95% CI) are represented in the figures. MS step estimates, at both wear locations, showed a significant increase across all sessions ($p < .05$). In contrast, AG step estimates, at both wear locations, were similar during SW and SJ sessions, and both were significantly greater compared with the SS session ($p < .05$). For kCal estimates, hip-worn MS and both AG devices showed a significant increase across all sessions ($p < .05$). Wrist-worn MS kCal estimates were similar for SS and SW, and both were significantly less than SJ ($p < .05$). MS points and AG

counts, at both wear locations showed a significant increase across all sessions ($p < .05$).

Figure 2 shows mean absolute differences (95% C.I.) between MS and DO counted steps, for each wear location (left axis). The right axis shows percent difference between MS estimates and DO steps. Compared with DO, the hip-worn MS significantly underestimated steps by 3.5% ($p < .05$) while the wrist-worn MS significantly overestimated steps by 4.2% ($p < .001$).

In FL conditions, participants engaged in 336 and 186 minutes (AG estimates) of MVPA during ACT and INACT, respectively. Figure 3 shows the estimated means (95% CI) from hip and dominant wrist-worn MS and AG PA outcomes during ACT and INACT weeks. Differences between ACT and INACT weeks for all MS and AG PA outcomes were significant using a linear mixed effects model ($p < .0001$).

Discussion

The primary aim of this study was to determine the sensitivity of the MS to detect changes in steps, points, and kCals during

Figure 1 — Mean (95% CI) for MS and AG PA outcomes across laboratory sessions. Mean (95% C.I.) for hip and dominant wrist-worn MS and AG PA outcomes across LAB sessions. *N* values for SS, SW, and SJ (respectively) are as follows: 1) AG and MS hip steps: 21, 21, 20; 2) AG and MS dominant wrist steps: 21, 21, 20; 3) AG hip kCals: 20, 20, 19; 4) MS hip kCals: 21, 21, 20; 5) AG dominant wrist kCals: 20, 20, 19; 6) MS dominant wrist kCals: 21, 21, 20; 7) AG hip counts and MS hip points: 21, 21, 20; 8) AG dominant wrist counts and MS dominant wrist points: 21, 21, 20. *Denotes significant increase from SS to SJ for given MS PA outcome ($p < .05$); •Denotes significant increase from SS to SJ for given AG PA outcome ($p < .05$); >Denotes SS not different than SW, both different than SJ for given MS PA outcome ($p < .05$); •Denotes significant increase from SS to SJ for given AG PA outcome ($p < .05$). Abbreviations: MS = Misfit Shine, AG = ActiGraph, PA = Physical Activity, SS = Sedentary; SW = Sedentary Plus Walking, SJ = Sedentary Plus Jogging.

Figure 2 — Mean differences (95% CI) between Misfit Shine (MS) and direct observation (DO) steps for laboratory condition. Right axis shows percentage difference between MS estimates and DO steps.

1) a structured laboratory-based protocol and 2) a two-week free-living protocol. This is the first study to report on the sensitivity of a consumer activity tracker to detect changes in PA. We found that when worn at the hip, MS estimates of steps, kCals, and points are sensitive to change in PA during LAB and FL conditions. When worn at the dominant wrist, MS estimates of steps and points were sensitive to change in activity during LAB sessions and FL conditions. MS wrist estimates of kCals were sensitive to differences between the ACT and INACT FL conditions, however they were not sensitive to differences between SS and SW LAB sessions.

In addition to taking a novel approach for examining sensitivity to detect change in PA, we sought to be consistent with the best practices for accelerometer validation by employing DO as the criterion measure of steps during the LAB sessions (Bassett et al., 2012). Compared with DO, the hip-worn MS significantly underestimated steps by 3.5%, while the wrist-worn MS significantly overestimated steps by 4.2%. Our study is the first to compare MS steps with DO as a criterion measure of behavior. Previous work compared MS step estimates to the optical detection system (OptoGait System OPTOGait, Microgate S.r.l, Italy 2010) during 30 minutes of treadmill walking at $4.8 \text{ km} \cdot \text{hr}^{-1}$. Compared with the OptoGait, MS steps were not significantly different (mean absolute percentage difference of 0.2%) (Kooiman et al., 2015). Our results

Figure 3 — Mean (95% C.I.) For MS and AG PA outcomes During ACT and INACT. Mean (95% C.I.) for hip and dominant-wrist worn MS and AG PA outcomes during ACT and SED. Differences between ACT and INACT for all PA outcomes recorded from both devices are significant ($p < .0001$). Reference: MS hip steps were 5,687 and 8,335 for INACT and ACT, respectively. MS wrist steps were 6,604 and 9,549 for ACT and INACT, respectively. Abbreviations: ACT; Active week, INACT; Inactive week. Abbreviations: MS = Misfit Shine, AG = ActiGraph, ACT = Active Week, INACT = Inactive Week, PA = Physical Activity.

suggest a greater difference between MS steps and the criterion during treadmill activity. This may be due to the differences in criterion measures used. We used a DO protocol which had two independent researchers count each step from video recording using the Observer XT software (Noldus). DO has been shown to be a valid criterion measure for estimating PA and SB (Lyden, Petruski, Staudenmayer, & Freedson, 2014). For the OptoGait there is no literature about the validity of this system in detecting steps. This system has only been validated against the Zebris gait analysis system (Zebris Medical GmbH, FDM-T system, Isny, Germany) for measuring kinematic parameters (speed, cadence, swing time, etc.). Secondly, in the study by Kooiman, participants wore the MS in their front pocket, while in our study it was worn on an elastic strap on the hip, and on the dominant wrist. Findings from both studies suggest the MS produces more accurate estimates of steps when worn on the hip or in the front pocket, compared with the wrist location.

The goal of the current study was to expand on previous research that has tested the validity of the MS (Bai et al., 2016; Brooke et al., 2017; Ferguson, Rowlands, Olds, & Maher, 2015). During a semi-structured lab protocol examining the validity of five wrist-worn consumer activity trackers the MS demonstrated the lowest accuracy and precision for estimated kCals, compared with indirect calorimetry (Bai et al., 2016). This laboratory protocol consisted of three sessions: 1) 20 minutes of sedentary activity, 2) 25 minutes of aerobic treadmill exercise at self-selected speeds, and 3) 25 minutes of resistance training, with 5 minutes of rest between sessions.

Participants wore the MS, and two other consumer activity trackers, on the left wrist, regardless of handedness. These results underscore the importance of correctly placing devices where the manufacturer recommends, and call attention to issues that might arise when placing multiple devices at one wear location. This study did not report how the three left wrist-worn devices were positioned relative to one another. Future work with consumer activity trackers should report these methodological details. Other groups have reported the convergent validity of the MS in FL environments. During a 48-hour FL protocol, kCal estimates from a wrist-worn MS were significantly correlated with estimates from a SenseWear Armband Mini. Mean absolute percentage error between MS and SenseWear Armband Mini estimates of kCals was 15.2% (Brooke et al., 2017). These findings are consistent with a similar study that also examined the convergent validity of a left wrist-worn MS during 48 hours of FL protocol. Compared with BodyMedia SenseWear Model MF, the MS underestimated kCals (bias: -479 kCal/day), but showed bias comparable to the FitBit One and FitBit zip (Ferguson et al., 2015). Criterion and convergent validity results for the MS are mixed but suggest that shorter duration bouts and rest periods during laboratory protocols may be problematic for the MS, while results from longer duration FL studies indicate acceptable validity. Our results contribute additional evidence suggesting the sensitivity of the MS to detect change.

The AG physical activity outcomes are not criterion measures of physical activity, however AG data from the present study are useful as comparison measures. For example, consider the change in hip-worn MS points and AG counts among LAB sessions. Although these outcomes have different scales, their respective y-axes (Figure 1) are relative to the lower and upper CI bounds for SS and SJ, respectively. Therefore, when considering MS points and AG counts it can be seen that the relative change among conditions is similar for both devices. This approach for visually

inspecting the similarities and/or differences between MS and AG PA outcomes may be useful for researchers trying to gauge the utility of other consumer activity trackers, compared with a research-grade accelerometer.

In our study, several factors affected our ability to perform a formal analysis of concurrent validity between MS and AG PA outcomes. First, a direct comparison between MS points and AG counts was not possible since the outputs are on different scales and the MS points and AG counts are computed with proprietary algorithms. These values should be positively associated but values are not directly comparable. Second, we elected not to directly compare MS and AG step estimates during the LAB condition because comparisons of MS and DO steps (criterion measure) are considered best practices for device validation (Bassett et al., 2012).

Compared with the AG, we found that the hip-worn MS significantly overestimated kCals for SW and SJ LAB sessions, while the dominant wrist-worn MS overestimated kCals during SJ. Although our results suggest that the MS tends to overestimate kCals during treadmill activity, previous work has shown that the Freedson equation (used to calculate AG kCals in the present study) underestimates kCals during similar activities (Lyden, Kozey, Staudenmayer, & Freedson, 2011). We acknowledge that using the Freedson equation to calculate kCals from the wrist-worn AG is a limitation of our study, however there are currently limited methods available for these estimations. Other groups have studied the relationship between MS estimates of kCals and criterion measures such as indirect calorimetry and doubly labeled water (DLW). One study compared MS kCal estimates to two criterion measures in two separate conditions. For condition one, MS kCals were compared with whole room calorimeter estimates during a 24-hour simulated free-living day. The room calorimeter protocol called for participants to do desk work, watch TV, housework, treadmill walking, and sleep. For condition two, MS estimates were compared with DLW collected over a 15-day free-living period. For the simulated free-living condition, MS kCal estimates were not significantly different compared with the room calorimeter kCals. However, for the free-living condition, MS kCal estimates were significantly lower than DLW kCals. The authors note that MS underestimations during the free-living condition were likely attributed to periods of device non-wear time (Murakami et al., 2016). If participants took the device off to bathe the device will only register kCal values representative of resting metabolic rate (RMR), while DLW will capture kCal values representative of the movement while bathing.

Despite not reporting time-stamped data, the MS has utility as a PA exposure or outcome measure for free-living individuals. During the FL condition of the present study, all PA outcomes from both MS devices (hip and dominant wrist) could differentiate between one week when participants were active and one week when they were less active. During the ACT and INACT weeks participants achieved 336 and 186 minutes of MVPA, respectively. Regardless of the fact that our participants met PA guidelines during the INACT week, our FL results suggest that the MS can detect a change of 150 minutes of MVPA (336–186 mins MVPA). These findings may be of particular interest to groups who are considering using an activity tracker as an intervention tool, as these devices are increasingly being used in such a capacity (Jakicic et al., 2016).

To our knowledge, this is the first study to use a comprehensive approach for testing a consumer activity tracker's sensitivity to change in PA. Our protocol included LAB and FL conditions, and we also report AG PA outcomes and comparisons with

DO-counted steps. Including the LAB condition was an important first step which provided highly controlled conditions where we knew exactly what participants were doing. This step is recommended when examining detection of change for other consumer activity trackers to ensure differences are in fact discernable. Future studies should also focus on examining device sensitivity to smaller empirical changes in PA. Given the public health importance of promoting PA, and rapid proliferation of consumer activity trackers, we expect more studies to use these technologies in research applications. Our findings expand on previous validation studies and suggest that the MS is a valid tool for estimating PA exposure in PA interventions and as an outcome to detect changes in PA.

Acknowledgments

This research was funded by the University of Massachusetts Amherst, Institute of Applied Life Sciences (IALS) seed grant initiative and matching funds from Fossil Group (Richardson, TX, USA).

References

- ActiGraph. (2012). GT3X+ and wGT3X+ device manual. Retrieved from https://dl.theactigraph.com/GT3Xp_wGT3Xp_Device_Manual.pdf
- Bai, Y., Welk, G.J., Nam, Y.H., Lee, J.A., Lee, J.M., Kim, Y., . . . Dixon, P.M. (2016). Comparison of consumer and research monitors under semistructured settings. *Medicine & Science in Sports & Exercise*, 48(1), 151–158. PubMed doi:10.1249/MSS.0000000000000727
- Bassett, D.R., Jr., Rowlands, A., & Trost, S.G. (2012). Calibration and validation of wearable monitors. *Medicine & Science in Sports & Exercise*, 44(1 Suppl. 1), S32–S38. PubMed doi:10.1249/MSS.0b013e3182399cf7
- Brooke, S.M., An, H.S., Kang, S.K., Noble, J.M., Berg, K.E., & Lee, J.M. (2017). Concurrent validity of wearable activity trackers under free-living conditions. *Journal of Strength & Conditioning Research*, 31(4), 1097–1106. PubMed doi:10.1519/JSC.0000000000001571
- Evenson, K.R., Goto, M.M., & Furburg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 159. PubMed doi:10.1186/s12966-015-0314-1
- Ferguson, T., Rowlands, A.V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 42. PubMed doi:10.1186/s12966-015-0201-9
- Fokkema, T., Kooiman, T.J., Krijnen, W.P., CP, V.D.S., & M, D.E.G. (2017). Reliability and validity of ten consumer activity trackers depend on walking speed. *Medicine & Science in Sports & Exercise*, 49(4), 793–800. PubMed doi:10.1249/MSS.0000000000001146
- Freedson, P.S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, 30(5), 777–781. PubMed doi:10.1097/00005768-199805000-00021
- Future Market Insights. (2017). Wearable fitness trackers market: Wrist wear product type segment expected to remain dominant over the forecast period: Global industry analysis (2012–2016) and opportunity assessment (2017–2027). Retrieved from <https://www.futuremarketinsights.com/reports/wearable-fitness-trackers-market>
- Greer, A.E., Sui, X., Maslow, A.L., Greer, B.K., & Blair, S.N. (2015). The effects of sedentary behavior on metabolic syndrome independent of physical activity and cardiorespiratory fitness. *Journal of Physical Activity & Health*, 12(1), 68–73. PubMed doi:10.1123/jpah.2013-0186
- 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 doi:10.1016/j.pcad.2016.02.006
- Hickey, A.M., Mendoza, A., & Freedson, P.S. (2015). Examination of the effect of monitor location and reliability of a consumer activity tracker. Paper presented at the New England American College of Sports Medicine, Providence, RI.
- Jakicic, J.M., Davis, K.K., Rogers, R.J., King, W.C., Marcus, M.D., Helsel, D., . . . Belle, S.H. (2016). Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: The IDEA randomized clinical trial. *Journal of the American Medical Association*, 316(11), 1161–1171. PubMed doi:10.1001/jama.2016.12858
- Kooiman, T.J., Dontje, M.L., Sprenger, S.R., Krijnen, W.P., van der Schans, C.P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, 7, 24. PubMed doi:10.1186/s13102-015-0018-5
- Lee, J.A., Williams, S.M., Brown, D.D., & Laurson, K.R. (2015). Concurrent validation of the Actigraph gt3x+, polar active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *Journal of Sports Sciences*, 33(10), 991–1000. PubMed doi:10.1080/02640414.2014.981848
- Lee, W.Y., Clark, B.K., Winkler, E., Eakin, E.G., & Reeves, M.M. (2015). Responsiveness to change of self-report and device-based physical activity measures in the living well with diabetes trial. *Journal of Physical Activity & Health*, 12(8), 1082–1087. PubMed doi:10.1123/jpah.2013-0265
- Lyden, K., Kozey, S.L., Staudenmeyer, J.W., & Freedson, P.S. (2011). A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European Journal of Applied Physiology*, 111(2), 187–201. PubMed doi:10.1007/s00421-010-1639-8
- Lyden, K., Petruski, N., Staudenmayer, J., & Freedson, P. (2014). Direct observation is a valid criterion for estimating physical activity and sedentary behavior. *Journal of Physical Activity & Health*, 11(4), 860–863. PubMed doi:10.1123/jpah.2012-0290
- Maher, C.A., Mire, E., Harrington, D.M., Staiano, A.E., & Katzmarzyk, P.T. (2013). The independent and combined associations of physical activity and sedentary behavior with obesity in adults: NHANES 2003–06. *Obesity (Silver Spring)*, 21(12), E730–E737. doi:10.1002/oby.20430
- 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 doi:10.1093/aje/kwm390
- Matthews, C.E., Keadle, S.K., Troiano, R.P., Kahle, L., Koster, A., Brychta, R., . . . Berrigan, D. (2016). Accelerometer-measured dose-response for physical activity, sedentary time, and mortality in US adults. *The American Journal of Clinical Nutrition*, 104(5), 1424–1432. PubMed doi:10.3945/ajcn.116.135129
- Mendoza, A.R., Hickey, A.M., & Freedson, P.S. (2015). A comparison of the misfit shine and ActiGraph GT3X + accelerometer in estimating steps during physical activity. Paper presented at the New England American College of Sports Medicine, Providence, RI.
- Montoye, A.H., Pfeiffer, K.A., Sutton, D., & Trost, S.G. (2014). Evaluating the responsiveness of accelerometry to detect change in physical

- activity. *Measurement in Physical Education and Exercise Science*, 18(4), 273–285. PubMed doi:10.1080/1091367X.2014.942454
- Murakami, H., Kawakami, R., Nakae, S., Nakata, Y., Ishikawa-Takata, K., Tanaka, S., & Miyachi, M. (2016). Accuracy of wearable devices for estimating total energy expenditure: Comparison with metabolic chamber and doubly labeled water method. *JAMA Internal Medicine*, 176(5), 702–703. PubMed doi:10.1001/jamainternmed.2016.0152
- Office of Disease Prevention and Health Promotion. (2008). Physical activity guidelines for Americans. Retrieved from <https://health.gov/paguidelines/guidelines/>
- R: A language and environment for statistical computing. (2017). R Foundation for statistical computing. Retrieved from <https://www.R-project.org/>
- Stables, J. (2018). Best fitness tracker guide 2018: The top activity bands you can buy now. Retrieved from <https://www.wareable.com/fitness-trackers/the-best-fitness-tracker>
- Swartz, A.M., Rote, A.E., Cho, Y.I., Welch, W.A., & Strath, S.J. (2014). Responsiveness of motion sensors to detect change in sedentary and physical activity behaviour. *British Journal of Sports Medicine*, 48(13), 1043–1047. PubMed doi:10.1136/bjsports-2014-093520
- Thompson, W.R. (2015). Worldwide survey of fitness trends for 2016: 10th Anniversary edition. *ACSMs Health & Fitness Journal*, 19(6), 9–18. doi:10.1249/FIT.0000000000000164
- Troiano, R.P., Berrigan, D., Dodd, K.W., Masse, L.C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, 40(1), 181–188. PubMed doi:10.1249/mss.0b013e31815a51b3
- Troiano, R.P., McClain, J.J., Brychta, R.J., & Chen, K.Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, 48(13), 1019–1023. PubMed doi:10.1136/bjsports-2014-093546
- Wright, S.P., Hall Brown, T.S., Collier, S.R., & Sandberg, K. (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 312(3), R358–R367. doi:10.1152/ajpregu.00349.2016