ABSTRACT

VALIDITY ASSESSMENT OF BIOMETRICS FOR THE FIRST-GENERATION APPLE WATCH AND MICROSOFT BAND DURING STEADY-STATE EXERCISE

William C. Kalamaras, M.S. Ed.
Department of Kinesiology and Physical Education
Northern Illinois University, 2016
Dr. Steve Howell, Co-Director
Dr. Amanda Salacinski, Co-Director

In recent years, wearable technology equipped with advanced biometric sensors has grown in both market share and consumer popularity. With these devices depending heavily upon the use of algorithms to estimate physiological parameters, the validation of biometric readings from these devices is paramount in establishing the credibility necessary to substitute for conventionally used equipment.

PURPOSE: To assess the biometric validity and reliability of heart rate (HR) and energy expenditure (EE) recorded by the first-generation Apple Watch (42mm) and Microsoft Band during submaximal aerobic exercise on a treadmill at steady state.

METHODS: Twenty-three participants (16 males, 7 females; mean age: 22.7 ± 3.8 years old) participated in three, 20-minute testing trials performing submaximal aerobic exercise: two trials assessed biometric validity for HR and EE in each wearable smartwatch (at 50% and 75% VO₂ max). A third testing trial (repeat of 75% VO₂ max; alternate wrist placement) assessed device reliability agreement across left and right wrists. Criterion values were obtained for HR (bpm) and EE (kCals) using a Polar H7 heart rate monitor (HRM) (Polar Electro Oy, Kempele, FIN) and ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT), respectively. Mean differences were assessed through a repeated-measures analysis of variance (ANOVA)
(α=0.05) for HR and EE independently. Follow-up paired t tests were run after significant main effects occurred. Pearson correlation coefficients (r) and standard error of estimates (SEE) were used to establish validity of each biometric, while intraclass correlation coefficients (ICC) based on a two-way random effects ANOVA model and an absolute agreement definition [ICC(2,1)] were used to assess device reliability agreement across left and right wrists.

RESULTS: Mean HR (bpm) values for testing trials one (50% VO₂ max) and two (75% VO₂ max) were recorded for their respective intensities at 145.2 ± 8.4 and 172.5 ± 9.3 bpm for the Apple Watch; 141.1 ± 8.0 and 165.1 ± 8.6 bpm for the Microsoft Band; and 145.1 ± 8.4 and 172.3 ± 9.5 for the Polar H7 HRM. Criterion comparisons revealed no significant difference between devices for Apple Watch HR (p = .164), and a 5.6 bpm difference (lower) for the Microsoft Band (p < .0005). Mean EE (kCals) values for testing trials one (50% VO₂ max) and two (75% VO₂ max) were measured for their respective intensities at 160.1 ± 51.1 and 212.2 ± 55.9 kCals for the Apple Watch; 182.7 ± 55.9 and 264.4 ± 55.0 kCals for the Microsoft Band; and 164.2 ± 51.2 and 221.1 ± 62.0 for the ParvoMedics 2400 TrueOne Metabolic System. Criterion comparisons revealed a 6.5 kCal difference (lower) for the Apple Watch (p = .002) and a 22.0 kCal difference (higher) for Microsoft Band (p =.03).

Pearson correlation coefficients (r) along with 95% confidence intervals (CI) and standard error of estimates (SEE, bpm) for HR at the respective intensities (50%, 75% VO₂ max) were computed at r=1.00 (1.00 – 1.00), (SEE=0.54) and r=0.99 (0.99 – 1.00), (SEE=1.04) for the Apple Watch and r=0.91 (0.80 – 0.96), (SEE=3.57) and r=0.90 (0.71 – 0.94), (SEE=4.83) for the Microsoft Band. Correlation coefficients and SEE (kCals) for EE were computed at r=0.98 (0.95 – 0.99), (SEE=10.07) and r=0.98 (0.95 – 0.99), (SEE=12.34) for the Apple Watch.
and \( r=0.49 \ (0.10 - 0.75) \), (SEE=45.59) and \( r=0.74 \ (0.49 - 0.88) \), (SEE=37.60) for the Microsoft Band.

Mean HR across wrists did not differ in the reliability agreement trials (repeated 75% \( \text{VO}_{2\text{max}} \) trial) for either the Apple Watch \((p = .18)\) or Microsoft Band \((p = .75)\). Mean EE between wrists also did not differ for either the Apple Watch \((p = .88)\) or Microsoft Band \((p = .80)\). ICCs for HR reliability agreement along with 95% CIs were 0.61 (0.28 – 0.81) for the Apple Watch and 0.51 (0.13 – 0.76) for the Microsoft Band; ICCs for EE reliability agreement (within a 95% CI) were 0.98 (0.96 – 0.99) for the Apple Watch and 0.73 (0.46 – 0.88) for the Microsoft Band.

**CONCLUSION**: The Apple Watch provides valid measurements of both HR and EE biometrics during both submaximal aerobic exercise intensities at steady state (50% and 75% \( \text{VO}_{2\text{max}} \)). While evidence for the Apple Watch does not establish reliability agreement across left and right wrists for HR, it does establish excellent reliability agreement for EE.

The Microsoft Band provides less valid measures of HR, which are more valid at a lower exercise intensity (50% \( \text{VO}_{2\text{max}} \)). The Microsoft Band does not provide valid measures of EE, nor does it provide reliable measures across left and right wrists for either HR or EE.

*Keywords*: biometrics; smartwatch; wearable technology; validity
VALIDITY ASSESSMENT OF BIOMETRICS FOR THE FIRST-GENERATION
APPLE WATCH AND MICROSOFT BAND DURING
STEADY-STATE EXERCISE

BY

WILLIAM CONSTANTINE KALAMARAS
©2016 William Kalamaras

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF SCIENCE IN EDUCATION

DEPARTMENT OF KINESIOLOGY AND PHYSICAL EDUCATION

Thesis Co-Directors:
Dr. Steve Howell
Dr. Amanda Salacinski
ACKNOWLEDGEMENTS

I would like to thank everyone who helped make this research project a reality. Thank you to those who volunteered their time and effort by participating, as well as to the numerous peers and faculty members from the Department of Kinesiology and Physical Education (KNPE). This project would not have been possible without your assistance. I am very grateful for all of my committee members; your time, support, and dedication throughout this process have been invaluable to me. I would like to give a special thanks to Brett Coryell and Dr. Stephen Haliczer from the Division of Information Technology (DoIT) for providing me with the wearable devices researched in this project. Without your support, this project would not have happened.

I would also like to express my deepest appreciation toward two committee members who were indispensable to me throughout this adventure. Thank you, Dr. Marilyn Looney, for your patience, moral support, and research expertise. I cannot thank you enough for the countless hours you spent working with me (at some of the most inopportune times). Finally, thank you, Dr. Amanda Salacinski, for serving as my mentor throughout this process. In spite of relocating from Northern Illinois University, I am extremely appreciative of the selflessness and generosity you displayed on a daily basis. I hold your expertise, devotion, and support as both my instructor and thesis adviser in the highest esteem possible.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Procedures</td>
<td>21</td>
</tr>
<tr>
<td>Pretesting Dietary Guidelines</td>
<td>21</td>
</tr>
<tr>
<td>( \text{VO}_2^\text{max} ) Assessment (Pretest)</td>
<td>22</td>
</tr>
<tr>
<td>Testing Sessions One and Two (T(_1) and T(_2))</td>
<td>24</td>
</tr>
<tr>
<td>Testing Session Three (T(_3))</td>
<td>25</td>
</tr>
<tr>
<td>Posttesting Procedures</td>
<td>25</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>26</td>
</tr>
<tr>
<td>RESULTS</td>
<td>27</td>
</tr>
<tr>
<td>Compliance</td>
<td>27</td>
</tr>
<tr>
<td>Participant Characteristics</td>
<td>27</td>
</tr>
<tr>
<td>Heart Rate (HR)</td>
<td>28</td>
</tr>
<tr>
<td>Energy Expenditure (EE)</td>
<td>31</td>
</tr>
<tr>
<td>Measures of Validity and Reliability</td>
<td>32</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>35</td>
</tr>
<tr>
<td>Limitations</td>
<td>40</td>
</tr>
<tr>
<td>Conclusion</td>
<td>42</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>44</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>46</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participant Descriptive Statistics</td>
<td>28</td>
</tr>
<tr>
<td>2. Mean Heart Rate at 50% and 75% VO$<em>2$$</em>{max}$</td>
<td>29</td>
</tr>
<tr>
<td>3. Mean Energy Expenditure at 50% and 75% VO$<em>2$$</em>{max}$</td>
<td>32</td>
</tr>
<tr>
<td>4. Pearson Correlation Coefficients ($r$) with 95% Confidence Intervals (CI) and Standard Error of Estimates (SEE) for Heart Rate and Energy Expenditure at 50% and 75% VO$<em>2$$</em>{max}$</td>
<td>33</td>
</tr>
<tr>
<td>5. Heart Rate Intraclass Correlation Coefficients (ICC) for 75% VO$<em>2$$</em>{max}$</td>
<td>34</td>
</tr>
<tr>
<td>6. Energy Expenditure Intraclass Correlation Coefficients (ICC) for 75% VO$<em>2$$</em>{max}$</td>
<td>34</td>
</tr>
<tr>
<td>7. Participant Descriptive Statistics by Session</td>
<td>83</td>
</tr>
<tr>
<td>8. Participant Descriptive Statistics by Session (Overview)</td>
<td>84</td>
</tr>
<tr>
<td>9. Concordance Correlation Coefficients ($r_c$) with 95% Confidence Intervals (CI) for Heart Rate and Energy Expenditure at 50% and 75% VO$<em>2$$</em>{max}$</td>
<td>86</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1. Schematized Overview of Experimental Procedures</td>
<td>13</td>
</tr>
<tr>
<td>2. Device by Intensity Interaction for Mean Heart Rate</td>
<td>30</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. REVIEW OF LITERATURE</td>
<td>46</td>
</tr>
<tr>
<td>B. INFORMED CONSENT FORM</td>
<td>62</td>
</tr>
<tr>
<td>C. PAR-Q</td>
<td>66</td>
</tr>
<tr>
<td>D. MEDICAL HISTORY QUESTIONNAIRE</td>
<td>68</td>
</tr>
<tr>
<td>E. RECRUITMENT SCRIPT</td>
<td>71</td>
</tr>
<tr>
<td>F. RECRUITMENT FLYER</td>
<td>74</td>
</tr>
<tr>
<td>G. SUBJECT INFORMATION SHEET</td>
<td>76</td>
</tr>
<tr>
<td>H. V₂ MASK SIZING GAUGE</td>
<td>78</td>
</tr>
<tr>
<td>I. MICROSOFT BAND SIZING GUIDE</td>
<td>80</td>
</tr>
<tr>
<td>J. PARTICIPANT CHARACTERISTICS TABLES</td>
<td>82</td>
</tr>
<tr>
<td>K. CONCORDANCE CORRELATION COEFFICIENTS</td>
<td>85</td>
</tr>
</tbody>
</table>
INTRODUCTION

We live in a world where technology has become ubiquitous with virtually every aspect of life. Evident in everything from engaging in communication to performing surgical procedures halfway across the world, there is simply no arguing with the fact that technology has revolutionized the way we live. While technological enhancement may in fact boost productivity and efficiency in most of what we intend it to, as consumers we sometimes need to be reminded that the price of progress is not always without compromise. In particular, mobile technology has become infamous for its manufacturers’ marketing gimmicks and outrageous statistical claims. While this may be an effective model for driving smartphone sales and innovation, as mobile wearable technology evolves, manufacturers will likely require a more calculated marketing approach.

Wearable technology is still an infant in the mobile sector, and its development in recent years has undoubtedly helped fuel the growing popularity of mobile technology even further. As if mobile smartphone capabilities were not advanced enough, much of the wearable market has been geared at further enhancing and expanding their capabilities. While many of the current designs will likely be phased out as the industry is refined, smartwatches and biometric sensors are here to stay. As this technology continues to tap further into the use of biological sensors, the complexity of the technology will likely follow suit. Eventually, the popular trend of increasing screen sizes from year to year will become insignificant and the accuracy of the biometric data reporting will be what allows the wearable market to flourish.
With the 2015 release of the first-generation Apple Watch, consumers have become familiarized with the concept of smartwatches and their dual role as tethered smartphone companions and activity trackers. Though the tech giant was not the first to capitalize on this concept, a Strategy Analytics Report suggests that Apple was responsible for more than half of all smartwatch sales in April of 2016 (Raskind, 2016). After releasing their first wearable device last year (2015), Apple, Inc., already controls half of the wearable market share in the US and is listed as one of seven key players in the American wearable industry alongside Fitbit, Inc.; Google, Inc.; Jawbone, Inc.; Nike, Inc.; Pebble Technology Corp.; and Qualcomm, Inc. (Wearable Technology Market by Product, Application, Type, & Geography - Global Forecast to 2020, 2015). Apple has helped popularize health-tracking smartwatches, but they are not stopping with the development of the Apple Watch. Their HealthKit framework provides a platform for expansion using a growing number of third-party devices and software capable of interfacing and sharing data (Smith, 2015). This will help to drive future wearable development and solidify Apple’s future in the world of informatics and health-tracking technology (Smith, 2015).

While the global impact of the wearable industry is already quite significant, it extends far beyond the influence of Apple. As the new kid on the block, Apple is an accessory to an already booming marketplace dominated by the United States (Wearable Technology Market by Product, Application, Type, & Geography - Global Forecast to 2020, 2015). In 2014, America held the largest market share in the entire wearables industry (Wearable Technology Market by Product, Application, Type, & Geography - Global Forecast to 2020, 2015). Additionally, the wearable technology market is projected to surge to a value of $31.27 billion USD by the year 2020, with wearables exclusively categorized as wearable “fitness” technology accounting for just under half
of that figure (*Wearable Fitness Technology Market by Product, Category, Component - Global Forecast to 2020*, 2016). From these numbers, it is evident that these devices are selling, and selling rather well. Furthermore, the market for smartwatches with biometric capabilities will likely continue to be a major influence in the world of wearable tech as long as these devices function as advertised ("Wearables: Fad or the Future?" 2015).

With an increasing opportunity for potential, wearable technology is providing a hopeful foundation for the future development of advanced healthcare devices. As the wearable and biosensor sector continues to grow, medical-grade wearable devices are already in development (Bryson, 2009). For the current generation of wearables, however, the biometric feedback is somewhat limited and highly dependent on proprietary algorithms ("Wearables: Fad or the Future?" 2015). What’s more, manufacturers are under no obligation to disclose their algorithms to the general public. In other words, while consumers may demand accurate devices for monitoring their physical activity and vitals, current wearables rely heavily upon the use of algorithms that are both undisclosed and unsubstantiated ("Wearables: Fad or the Future?" 2015). As a result, consumers will have to rely on trial and error to gauge wearable biometric accuracy. With each manufacturer likely to claim superiority for their algorithms and biometric sensor accuracy, these devices will need to undergo experimental trials on a per-device basis in order to gain any valid insight about their performance.

Those familiar with the current generation of wearables may recall the not-so-distant past and the popularity of wrist-worn activity (or fitness) trackers. To name a few, some of the more popular trackers include the Fitbit line, the Jawbone Up series, BodyMedia bands, and Nike’s Fuel band. These trackers, which were essentially glorified accelerometer/gyroscope combinations,
have gradually evolved into today’s wearable smartwatch and share a number of similarities with them (Lee, Kim, & Welk, 2014). These fitness trackers helped evoke the current trend of wearing activity trackers at the wrist and, similar to today, relied on a combination of sensor biofeedback and algorithms to provide basic movement statistics (Lee et al., 2014). At the most basic level, older devices usually provided an assessment of step count (SC), distance traveled, energy expenditure (EE), time active, and (occasionally) sleeping habits (Lee et al., 2014). Perhaps most importantly, the use of algorithms which was commonplace in the development of early trackers, is still very much relevant in today’s wearable smartwatch market ("Wearables: Fad or the Future?" 2015).

When working with wearable sensors, caution should be exercised as to not generalize the accuracy of biometric tracking from one device to another. In one study that measured the accuracy of energy expenditure (EE) in a number of different activity trackers, this practice can be justified. Researchers from Illinois State University measured the percent error of EE for eight different wearables using indirect calorimetry and 60 males and females. During 69 minutes of physical activity, the percent error of EE in these devices was found to range between 9.3% and 23.5% (Lee et al., 2014). In order from most to least accurate, percent error for EE was 9.3% for the BodyMedia FIT armband, 10.1% for the Fitbit Zip, 10.4% for the Fitbit One, 12.2% for the Jawbone Up, 12.6% for the ActiGraph, 12.8% for DirectLife, 13.0% for the NikeFuel Band, and a whopping 23.5% for the more developed Basis B1 Band (Lee et al., 2014).

These results are just one demonstration of how various levels of biometric accuracy are obtained through the use of different wearable devices. While this substantiates the practice of validating biometrics on each device independently, these results also suggest that more accurate
measurements are attained with trackers worn closer to the core of the body, as opposed to being worn on the wrist (Lee et al., 2014). Given the alarming number of fitness trackers and smartwatches that are specifically designed to be worn on the wrist, it would seem that device engineers may be sacrificing accuracy of biometrics in order to capitalize on the popularity of the smartwatch market (Lee et al., 2014; "Wearables: Fad or the Future?" 2015). In any case, it is clear that these devices, even when produced by the same manufacturer, must undergo individualized testing for biometric validation.

Smartwatches and Activity Trackers

In previous-generation wearables, advanced algorithms were paramount in a device’s ability to produce meaningful health metrics (Kim, Wang, & Mahmud, 2016). Without these complex operations to interpret sensor inputs, a wrist-full of biosensors would provide no meaningful data. In other words, algorithms were the powerhouses that computed otherwise ambiguous numbers into meaningful metrics (such as SC, EE, and distance; Kim et al., 2016; Lee et al., 2014). In comparison to aging fitness trackers, the use of more sophisticated (or a greater number of) sensors found in current-generation wearables provides more biofeedback (Kim et al., 2016). As such, the Apple Watch and several other modern-day wearables are able to further refine biometric output (Dobkin & Dorsch, 2011). While more input theoretically allows for a more accurate result, it also means a greater ability to erroneously skew output (Papi, Osei-Kuffour, Chen & Mcgregor, 2015). In this regard, biometrics reported by newer smartwatches and activity trackers may be more sensitive to extraneous testing conditions and as a result may be more error prone. Accordingly, these algorithms grow in complexity and their potential to disproportionately impact the accuracy.
Much like the Apple Watch, the Microsoft Band is a first-generation smartwatch boasting similar functionality and biometric capability. As with many other smartwatches, it is designed to be worn at the wrist and features a particularly long list of biosensors ("Getting Started with Your Microsoft Band," n.d.). Though it was released in 2014, its feature set still rivals that of the year-older Apple Watch. Unlike the Apple Watch, however, the Microsoft Band incorporates a galvanic skin sensor, a global positioning system (GPS) sensor, and an ultraviolet (UV) light sensor ("Getting Started with Your Microsoft Band," n.d.). When it was first introduced in 2014, the original Microsoft Band was one of the most advanced smartwatches and activity trackers on the market. Now, within a few years of its release, its spec set has made an impression in the wearable smartwatch world as other device makers are increasingly following suit.

Devices such as the Microsoft Band and Apple Watch, while similar on paper, present great examples of devices that may inappropriately be generalized (in terms of accuracy). At first glance, both premium-grade wearables share many common core sensors and specifications. This may mislead the consumer into believing that specifications parallel performance and accuracy. We now know better, however, that the unique engineering and algorithms of these devices create entirely unique smartwatches which cannot be generalized to other wearables (Papi et al., 2015). Until there is an industry standard for the algorithms and sensors used in these devices, each device must undergo testing to appropriately gauge biometric accuracy. In particular, when dealing with data that is often reflective of physiological processes, these biomarkers can have critical implications for an individual’s overall health or state of well-being.

If it was not yet apparent, biomedical informatics and the accessibility of accurate biometric markers from all wearables is pivotal to the innovation and development of future
wearable technology. Apple and Microsoft have both demonstrated their understanding for the momentous role biometrics from these devices will have in the science community and for the Internet of Medical Things (IoMT). Through the use of their ResearchKit, Apple has equipped its users with the ability to channel large amounts of biometric data to health/medical research projects (Kim et al., 2016). Similarly, Microsoft actively develops their HealthVault platform, an ecosystem for interconnecting health and fitness devices and streamlining biometric health data. Apple and Microsoft are also among the biggest investors in the wearable biosensor market (Wearable Fitness Technology Market by Product, Category, Component - Global Forecast to 2020, 2016). While these projects will undeniably benefit future wearable and IoT development, they still do not help establish any level of biometric validity in either of the corporation’s flagship smartwatches.

These two wearables may not have been intended to directly compete with one another, but they warrant mention because of their similarity in functionality. Both devices are engineered by respectable corporations that are actively engaged in the future of wearables and biosensor research (Wearable Fitness Technology Market by Product, Category, Component - Global Forecast to 2020, 2016). Because these devices are so similar in capability, yet different enough in engineering and structural design, an assessment of biometric validity for both devices may offer insight for the direction of future smartwatch development and design. Furthermore, because it is currently unknown whether most smartwatches validly measure physiological metrics, their validation through a set of practical conditions is a necessary step to better understand the accuracy of metrics in today’s wearables (Dobkin & Dorsch, 2011).
Purpose

The purpose of the research study was to assess the validity (accuracy) and reliability (across left and right wrists) of two significant physiological biometrics as reported by the first-generation Apple Watch (42mm; Apple, Inc., Cupertino, CA) and Microsoft Band (Microsoft, Inc., Redmond, WA) during submaximal, steady state exercise on a treadmill. The biometrics being analyzed in both devices include heart rate (HR, bpm) and energy expenditure (EE, kCals).
DATA AND METHODOLOGY

Experimental Overview

Twenty-six participants underwent a total of three testing trials performing steady-state aerobic exercise on a DESMO Woodway Treadmill (Woodway USA, Waukesha, WI). During each session, participants were equipped with two wearable smartwatches (simultaneously): the 42mm Apple Watch (Apple, Inc., Cupertino, CA) and Microsoft Band (Microsoft, Inc., Redmond, WA). In the first two testing trials, workload was held constant at 50% (trial one / T1) and 75% (trial two / T2), respectively, of maximal aerobic workload (VO$_2$ max) for 20 minutes. These two testing trials assessed the biometric validity of heart rate (HR) and energy expenditure (EE) in each wearable device. Validity was determined through data analyses which compared wearable HR (bpm) and EE (kCals) measurements to criterion measures of a Polar H7 heart rate monitor (HRM) (Polar Electro Oy, Kempele, FIN) and ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT), respectively. A third testing session (repeat of 75% VO$_2$ max, alternate wrist placement) was conducted to determine if either smartwatch could be reliably interchanged across left and right wrists.

Participants

Twenty-six participants were recruited for this research study. Of the 26 recruits, the 23 who completed all testing trials were used in data analyses. Participants consisted of males and females between the ages of 18 and 40 years old, and eligibility was dependent upon the
completion of a mandatory screening process. All requirements were to be satisfied as specified in screening subsection “Inclusion Criteria.” The “Exclusion Criteria” subsection specifies conditions used to establish ineligibility.

Screening

Everyone underwent a mandatory screening process to screen for eligibility. The following three documents were completed and furnished to the principal investigator during this process: an informed consent form (Appendix B) stating that the individual understood his/her rights, as well as any risks, requirements, and expectations associated with being a participant in this research study; a Physical Activity Readiness Questionnaire (PAR-Q; Appendix C) to determine whether the individual satisfied the minimum requirements to safely participate in physical activity (in general); and a Medical History Questionnaire (Appendix D) used to disclose any known allergies and prescription medications, as well as to screen for specific medical conditions, disease risk factors, and other miscellaneous/general criteria.

Inclusion Criteria

Eligibility was granted by an individual’s ability to satisfy (at minimum) these five specific criteria during screening: (a) the individual was between the ages of 18 and 40 years old; (b) the individual had completed and furnished original copies of all required screening forms (with a valid date and signature) to the principal investigator; (c) all responses on the PAR-Q form (Appendix C) were answered “no,” demonstrating the individual’s state of physical activity readiness; (d) responses on the Medical History Questionnaire (Appendix D) classified the
individual as having a low cardiovascular disease (CVD) risk as outlined by the American College of Sports Medicine (Pescatello, Arena, Ribe, & Thompson, 2014). This classification was satisfied by answering “no” to all items listed under Parts I (Known Diseases) and II (Signs and Symptoms) and by answering “yes” to no more than one item under Part III (Coronary Artery Disease Risk Factor) (Pescatello et al., 2014); and (e) the individual was free from protocol-specific conflicts that would otherwise contraindicate participation. These specific conditions are described in the following “Exclusion Criteria” subsection.

**Exclusion Criteria**

Responses on the Medical History Questionnaire (Appendix D) matching any of the following criteria were used to nullify an individual’s eligibility to participate. Using information obtained in Part IV (Musculoskeletal Conditions and Other): (a) Current or previous musculoskeletal injuries (within the last 12 months) were indicated, which may have impaired the individual from safely completing exercise and physical activity requirements. (b) The presence of any tattoo(s) on the posterior (dorsal) side of the wrist was indicated. Upon inspection, the tattoo(s) was/were confirmed to cover any area three inches proximal of dorsal tubercle of the radius and the styloid process of the ulna. This may have interfered with the transmission of either smartwatch’s optical heart rate sensor. Under Part V (Female Health): (c) A positive state of pregnancy or an active attempt to become pregnant was indicated, which may have unnecessarily compromised prenatal health during testing protocol (females only). Under Part VI (General Supplemental Health): (d) Any prescription medication(s) and/or medical condition(s) that may have negatively impacted the ability to safely participate in testing protocols was/were indicated.
These responses were evaluated on an individual basis by the principal investigator. (e) A known allergy to silicone or silicone-based material was indicated, which may have increased the risk of anaphylaxis while wearing the required $V_2$ oxygen mask during testing trials.

In addition to these conditions, any atypical conditions or set of circumstances which may have interfered with an individual’s ability to complete testing requirements or to safely participate in any part of the study for any reason were used to exclude that particular individual from becoming a participant. These cases were evaluated by the principal investigator on an individual basis.

Recruitment

Participants were recruited via posted flyers (Appendix F) within locations around NIU’s Anderson Hall, Recreation Center, and in some local area running clubs within a 50-mile radius of the university. Individuals expressing interest to participate in the research study were provided with a Recruitment Script (Appendix E) by the principal investigator (either via e-mail or in person) which gave an overview of the research process, relevant procedures, and instructions on how to begin the screening process. An initial pretesting appointment was then scheduled with the principal investigator for each individual to complete the screening process and ask questions.

Procedures

Standardized testing protocols were instated throughout the duration of the study. Detailed protocols were utilized for the maximal oxygen uptake assessment ($VO_{2\text{, max}}$ assessment), each testing session (3), and the posttesting period. Participants were also required to adhere to a set of pretesting dietary guidelines. Figure 1 schematizes main (general) points for every aspect of this
study; this includes the aforementioned participant screening and registration processes as well as all pre-, intra-, and post-testing experimental procedures. Detailed protocols for each procedural component are outlined in their respective subsections.

**Subject Recruitment**
- Eligibility screening
- Consent process
- Registration

**Pretesting Appointment**
- Equipment sizing
- Initial anthropometrics (height/weight)

**VO₂ max Assessment**
- Modified Balke treadmill VO₂ max protocol (~45 minutes total)

**Testing Sessions (3)**
- Twenty minutes of physical activity (each)
- Twenty minutes of preparation (each)
- 36-48 hours between each session
- Standardized procedures (summary in right box)
  1. Trial 1 (T₁): 50% VO₂ max Apple: L. Wrist, Micro: R. Wrist
  2. Trial 2 (T₂): 75% VO₂ max Apple: L. Wrist, Micro: R. Wrist
  3. Trial 3 (T₃): 75% VO₂ max Micro: L. Wrist, Apple: R. Wrist

**Procedural Overview**
1. Participant empties bladder (BIA requirement)
2. Equipment preparation (warm-up)
3. Equipment calibration
4. 15-minute standing period (pretesting guidelines)
5. Time synchronization (all electronic devices)
6. InBody520 BIA (record new anthropometrics)
7. Smartwatch calibration (use new anthropometrics)
8. ParvoMed 2400 / Polar H7 (prepare / equip)
9. Smartwatch integrity tests (prepare / equip)
10. Initialize testing session
   - Start exercise activities on smartwatches
   - Log start time (laboratory wall clock)
   - VO₂ / Aerobic W. Load: (50% T₁; 75% T₂/T₃)
     - Speed changes (gross changes)
     - Grade changes (fine change)
   - ≥ 15 minutes at sub-maximal steady-state
11. Terminate testing session (stop data collection)
12. Cool-down period (~2 minutes)
   - Participant HR < 150 bpm
   - Treadmill speed/grade: 2.5 mph at 0% grade
13. Data export (.csv/.xml) by uID & session identifier
14. Standard equipment cleaning
15. Factory reset of all applicable wearables

**Figure 1: Schematized Overview of Experimental Procedures**

Pretesting Procedures

If the individual’s eligibility was approved by the principal investigator using the specified criteria, the participant was assigned a unique identification code (uID). All approved participants were then required to schedule their VO₂ max pretest (~45-minutes) and three testing sessions (~35-
40 minutes each). Including the VO$_2$$_{\text{max}}$ assessment (pretest), sessions were scheduled over four one-hour testing blocks at least 36-48 hours apart from one another. The total time commitment for all four sessions was approximately 165 minutes (~3 hours), with approximately 80 minutes being spent exclusively on exercising (~1.5 hours).

**Anthropometrics**

Anthropometric measurements, in addition to gender and birth date (mm-dd-yy) were used to calibrate testing equipment during each session. The ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT), 42mm Apple Watch (Apple, Inc., Cupertino, CA), and Microsoft Band (Microsoft, Inc., Redmond, WA) all required these measurements. Participants had their anthropometric measurements taken immediately prior to both pretesting and all testing trials. These measurements were recorded for each participant on their Subject Information Sheet (Appendix G).

On the day of the VO$_2$$_{\text{max}}$ assessment (pretest), the participant emptied his/her bladder and removed their shoes. Body mass (kg) was recorded using a T500E-B digital scale (Totalcomp, Prospect Park, NJ), and height (cm) was recorded using a Model 220 stadiometer (SECA, Birmingham, GBR). Imperial measurements for body mass (lbs) and height (in) were calculated by multiplying metric units by 2.20462 (lbs ∙ kg$^{-1}$) and 0.39370 (in ∙ cm$^{-1}$), respectively. Gender and birth date (mm-dd-yy), which were obtained during the screening process, were transferred to the Subject Information Sheet (Appendix G) for each participant.

For subsequent testing sessions (3), the recorded values for birth date (mm-dd-yy), gender, and height (cm & in) were re-used. New measurements were taken for body mass (kg & lbs), BMI
(kg · (m$^2$)$^{-1}$), and cellular fluid content (%) on the day of each testing session using bioelectrical impedance analysis (BIA). All BIAs were assessed using an InBody520 (Biospace Inc., Los Angeles, CA). Measurements were taken prior to beginning exercise protocols but after 15 minutes of standing upright to ensure proper fluid balance. To maintain consistent BIA measurements, participants were instructed to adhere to a set of dietary restrictions on days of testing. These restrictions included maintaining consistent dietary habits on days of testing and avoiding the excessive consumption of food and drink (1.5 hours), caffeine (5 hours), and alcohol (12 hours) prior to their scheduled testing time. These guidelines were included as part of the Recruitment Script (Appendix E) that was provided to every participant during the recruitment/registration process.

*Equipment Sizing and Preparation*

Before any pretesting or testing sessions were conducted, each participant was measured and pre-fitted for all necessary equipment. Equipment size and approximate clasp position (for relevant equipment) was recorded for each participant on his or her Subject Information Sheet (Appendix G). All participants were fitted for a properly sized V$_2$ mask (and accompanying harness) assembly (Hans Rudolph, Shawnee, KS), Polar Heart Rate Soft Strap (Polar Electro Oy, Kemple, FIN), Microsoft Band (Microsoft, Inc., Redmond, WA), and Apple Watch Sport Band (Apple, Inc., Cupertino, CA). Sizing was assessed and checked using the standardized protocol and proper fit was re-checked before each testing session.

Proper V$_2$ mask size (Hans Rudolph, Shawnee, KS) was assessed using the V$_2$ mask sizing gauge (Appendix H). The top of the sizing gauge (cyan/teal colored tip) was first placed atop the
nasal bridge at the deepest depression of the nasal bone. It was then rotated inferiorly to the face (near the mouth), and the mask size of the assembly was determined by following the marking below the chin. From smallest to largest, mask sizes ranged from petite (P) to large (L). The harness size of the assembly was determined by the mask size; the small (S) harness was used for mask sizes P through S, and the medium (M) harness was used for mask sizes M and L. Due to the adjustable ranges, proper size of the Polar Heart Rate Soft Strap (Polar Electro Oy, Kempele, FIN) was determined through observation. Stature was used to estimate the best fit for one of two sizes, XS-S or M-XXL. The strap was then test fitted and securely tightened circumferentially under the participant’s chest using the sliding clasp. If the clasp was positioned towards the terminal end in either direction, the other size was tested, and the strap yielding the least amount of movement was recorded on the Subject Information Sheet (Appendix G).

For proper smartwatch sizing, the Microsoft Band sizing guide (Appendix I) was used. The participant’s wrist was placed within the reference lines, and the proper size for the Microsoft Band (Microsoft, Inc., Redmond, WA) was determined as the smallest size outlining the wrist. Available sizes were S, M, and L. Using the recorded Microsoft Band (Microsoft, Inc., Redmond, WA) size as a reference, the Sport Watch band size for the Apple Watch (Apple, Inc., Cupertino, CA) was then determined. For L Microsoft Band (Microsoft, Inc., Redmond, WA) measurements, the M/L Sport Watch band was used; the S/M band was used for S and M measurements. Both smartwatches were then test fitted (for each wrist). The Microsoft Band (Microsoft, Inc., Redmond, WA) was oriented with the optical heart rate sensor and closing clasp positioned on the posterior (dorsal) side of the wrist. The Apple Watch (Apple, Inc., Cupertino, CA) was oriented with the optical heart rate sensor on the posterior side of the wrist and the Sport Watch band clasp
on the anterior (ventral) side of the wrist. For both smartwatches, precise clasp positioning was determined by carefully tightening/loosening one notch at a time, then attempting to displace the device perpendicularly along the skin of the participant’s forearm. This process was repeated until either device no longer slid along the participant’s skin. If either device reached the terminal fastener, the next size up or down was tested. The size and fastener position yielding the least amount of movement was then recorded for each device (for each wrist).

On testing days, equipment in the proper sizes was prepared prior to the participant’s arrival. Devices were then equipped and tested for proper fit in the following manner. First, the V2 mask assembly (Hans Rudolph, Shawnee, KS) was placed around the participant’s oral/nasal cavities so that the mask completely sealed around the nose and chin. While the participant held the assembly in place, the investigator secured the mask using the accompanying harness. The participant was then instructed to inhale deeply and expire normally while the inlet/outlet ports are manually obstructed by the investigator. The positive pressure from expiration was used to detect the presence of leaks. If no leaks were present, the outlet port was then connected to a six-foot expiration hose. The six-foot hose was used to transmit expired gas from the participant to the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT) during testing. If leaks were present, the mask was reseated and the harness was tightened incrementally until no air escaped during expiration.

Next, the Polar Heart Rate Soft Strap (Polar Electro Oy, Kemple, FIN) was equipped. The two electrodes on the inner-most surface of the strap were moistened with water, and the measured strap size was affixed directly inferior to the participant’s pectoral muscles. Proper fit was checked by attempting to displace the backside of the soft strap inferiorly along the participant’s torso. The
clasp was tightened until sliding the back side of the strap yielded no effect on the positioning of the strap’s front side. Finally, the Polar H7 HRM module (Polar Electro Oy, Kemple, FIN) was connected to the receiver socket located on the front of the soft strap.

Both smartwatches were then affixed to the participant’s wrists using the recorded sizes and positions. Device placement varied depending on which test was being conducted. For testing sessions one and two (T₁ and T₂; 50% and 75% \( \text{VO}_2 \text{max} \)), the 42mm Apple Watch (Apple, Inc., Cupertino, CA) was affixed to the participant’s left wrist while the Microsoft Band (Microsoft, Inc., Redmond, WA) was affixed to the right wrist. For testing session three (T₃) (75% \( \text{VO}_2 \text{max} \)), device placement was alternated. During all tests, both smartwatches were orientated such that the optical heart rate sensor faced the posterior side of the wrist. Devices were carefully aligned proximally to the dorsal tubercle of the radius and the styloid process of the ulna for the respective wrist. Integrity was then checked by attempting to displace both devices perpendicularly along the skin of the forearm. If either device was easily displaced along the skin, that device was tightened to the next notch and then retested. The investigator then performed a secondary check for proper alignment/positioning in respect to radial and ulnar landmarks. Once both devices were confirmed to align properly, the equipment underwent subsequent calibration and setup procedures.

**Calibration**

Testing equipment underwent a stringent number of calibration procedures prior to each testing session. The ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT) was calibrated to specific participant and environmental conditions while both smartwatch-linked smartphones underwent timing synchronization and participant calibration using
anthropometrics data. Prior to the first participant’s arrival, the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT) was powered on and warmed up for a minimum of 20 minutes using a timer. During this period, other smartwatch calibration procedures were conducted. The input of anthropometric parameters for all devices was conducted after the participant had received their BIA for that testing session.

Once the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT) had warmed up for the recommended 20 minutes, the gas analyzer and flowmeter components of the system were calibrated. Ambient temperature (°C), relative humidity (RH) (%), and barometric pressure (mmHg) from a Vantage VUE Wireless Weather Station 6351 (Davis Instruments, Hayward, CA) was entered into the ParvoMedics TrueOne 2400 Metabolic System software (OUSW 4.3.4). Using the onscreen guide, an O2/N2 gaseous mixture and several atmospheric air samples were sent to the gas analyzer and flowmeter, respectively. If calibration changes greater than ± 1.0% for the gas analyzer or greater than ± 3.0% for the flowmeter were detected, they were rejected and procedures were repeated after a five-minute period. Both gas analyzer and flowmeter components were recalibrated after three hours or after every third consecutive testing session. Next, the software was calibrated to participant-specific variables. Gender, birth date (mm-dd-yy), height (cm), and weight (kg) were entered in order to calibrate the ParvoMedics TrueOne 2400 Metabolic System software (OUSW 4.3.4) to the participant. Participant calibration procedures were repeated after each testing session had been conducted.

Similar to the Metabolic software (OUSW 4.3.4), participants’ data were used to calibrate both smartwatches to each participant before testing. Participant gender, birth date (mm-dd-yy), height (in.), and weight (lbs) were entered into the Apple Health (iOS 9.2.1) and Microsoft Health
smartphone applications on an Apple iPhone 5s smartphone (Apple, Inc., Cupertino, CA) and Samsung Galaxy S5 smartphone (Samsung, Guangdong, CHN), respectively. This portion of calibration procedures was conducted only after the participants had completed their pretest BIA. Prior to entering new participant information, any outstanding testing data and participant parameters were wiped in accordance with standardized protocol.

Time synchronization was the final step of equipment calibration. This step was performed in order to ensure that time stamps on all testing data sheets were properly aligned during data analysis. This step was crucial because data were collected from three independently operating devices. These procedures were repeated after each testing session and double-checked prior to test initiation. Before each session, the investigator noted the current time (hh:mm:ss) displayed on the laboratory wall clock and, if necessary, manually synchronized all devices to match the laboratory wall clock (down to the exact second). Devices for this procedure included the computer hosting the ParvoMedics TrueOne 2400 Metabolic System software (OUSW 4.3.4), the Apple iPhone 5s smartphone (Apple, Inc., Cupertino, CA), and the Samsung Galaxy S5 smartphone (Samsung, Guangdong, CHN). Because smartwatch time was synchronized to its smartphone counterpart, time synchronization for the 42mm Apple Watch (watchOS 2.1) and Microsoft Band (FW 10.03.3304.0 09R) were accomplished by changing the native time on the iPhone 5s (iOS 9.2.1) and Galaxy S5 (Android 5.0.1), respectively.
Testing Procedures

Each participant underwent a mandatory VO\textsubscript{2\text{max}} assessment (pretest) and three 20-minute testing sessions on a DESMO Woodway Treadmill (Woodway USA, Waukesha, WI) at 50% (T\textsubscript{1}) or 75% (T\textsubscript{2} / T\textsubscript{3}) of their determined VO\textsubscript{2\text{max}} (maximal aerobic workload). All testing sessions were conducted in the Human Performance Lab located in 206 Anderson Hall at NIU. During the VO\textsubscript{2\text{max}} pretest, participants were equipped with a Polar H7 HRM (Polar Electro Oy, Kemple, FIN) and connected to the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT). During each subsequent testing session, participants were equipped with a Polar H7 HRM (Polar Electro Oy, Kemple, FIN), 42mm Apple Watch (Apple, Inc., Cupertino, CA), and Microsoft Band (Microsoft, Inc., Redmond, WA) and connected to the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT). The Apple iPhone 5s (Apple, Inc., Cupertino, CA) and Samsung Galaxy S5 smartphones (Samsung, Guangdong, CHN) were paired with the 42mm Apple Watch (Apple, Inc., Cupertino, CA) and Microsoft Band (Microsoft, Inc., Redmond, WA), respectively.

Pretesting Dietary Guidelines

In accordance with the “Anthropometrics” section and the Recruitment Script (Appendix E), participants were instructed to adhere to a set of dietary guidelines on days of pretesting and testing sessions. These guidelines were in place in order to provide consistent readings during BIAs on the InBody520 (Biospace Inc., Los Angeles, CA). They were also intended to help avoid temporary/artificial fluctuations in body mass (kg) and/or hydration status (ratio of extracellular: total body water [ECW:TBW]).
On testing days (not including the pretest day), all participants were required to remain standing for 15 minutes before they could proceed to the InBody520 (Biospace, Inc., Los Angeles, CA) for their BIA in order to ensure proper fluid balance. The new BIA report generated for each session was used in accordance with procedural guidelines outlined in the “Anthropometrics” section. After the participant was equipped with the proper equipment, anthropometric measurements were used to calibrate testing equipment, and the exercise component of the testing session was initiated.

\( \text{VO}_{2\text{max}} \) Assessment (Pretest)

Each participant was first required to perform a modified Balke treadmill test (~45-minutes total) to assess their aerobic endurance, or capacity for maximal oxygen uptake (\( \text{VO}_{2\text{max}} \)). Percentages of this pretest assessment were used to set submaximal workloads for each subsequent testing session (50% \( \text{VO}_{2\text{max}} \) for \( T_1 \); 75% \( \text{VO}_{2\text{max}} \) for \( T_2 / T_3 \)). To perform the modified Balke test, the principal investigator equipped the participant with the Polar H7 HRM (Polar Electro Oy, Kemple, FIN) and connected them to the ParvoMedics TrueOne 2400 Metabolic System (ParvoMedics, Salt Lake City, UT).

Once all equipment was properly equipped and calibrated, the principal investigator initiated a maximal exercise test within the Metabolic software (OUSW 4.3.4). Treadmill speed was increased to a speed between 3.4 – 3.8 mph in order to mimic a moderate walking pace and subsequently increased/decreased in three-minute intervals until the participant’s respiratory exchange ratio (RER) fell between a value of 0.85 - 0.90. Once these criteria were met, the incline of the treadmill was increased every two minutes in the following progression: 2%, 5%, 8%, 11%,
and finally 14% incline. If the participant reached the maximum 14% incline, they continued on
the treadmill until volitional fatigue occurred. In either case, the participant was instructed to
continue running on the treadmill until he or she could no longer continue (volitional fatigue). At
that point, the participant was instructed to straddle the treadmill until the speed/incline was slowed
to 2.5 mph and lowered to an incline of 0%. Using this speed/grade, they performed a two-minute
cool-down session. Once complete, the treadmill was stopped and the participant was seated on a
chair to safely remove the testing equipment.

\( \text{VO}_2 \text{max} \) (mL \cdot kg\(^{-1}\) \cdot min\(^{-1}\)) was determined using criteria in accordance with the ACSM's
exercise testing and prescription guidelines (Pescatello et al., 2014). These criteria include: \( a \) The
participant’s HR was within \( \pm 5 \) bpm of age-predicted \( \text{HR}_{\text{max}} \) \( (208 \ - (0.7 \cdot \text{Age}) \) (Pescatello et al.,
2014); \( b \) a leveling off of \( \text{VO}_2 \) (over a \( \sim 30\)-second period) was observed, indicating the inability
to take in a larger amount of oxygen; and \( c \) a respiratory exchange ratio (RER) \( \geq 1.1 \) was
observed. Once the participant’s \( \text{VO}_2 \text{max} \) value (mL \cdot kg\(^{-1}\) \cdot min\(^{-1}\)) had been determined, it was
recorded under the Maximal \( \text{VO}_2 \) Assessment / 100% condition on the Subject Information Sheet
(Appendix G); 50% and 75% of this \( \text{VO}_2 \text{max} \) value was calculated and recorded under conditions
\( T_1 \) and \( T_2/T_3 \), respectively. These submaximal intensities were used as steady-state targets for the
three testing trials. If the participant’s assessed \( \text{VO}_2 \text{max} \) was 50 mL \cdot kg\(^{-1}\) \cdot min\(^{-1}\), this meant that
the intensity of the remaining three sessions was to be set at the following intensities: testing
session one (50% \( \text{VO}_2 \text{max} \)): 25.0 mL \cdot kg\(^{-1}\) \cdot min\(^{-1}\), testing sessions two and three (75% \( \text{VO}_2 \text{max} \)): 37.50 mL \cdot kg\(^{-1}\) \cdot min\(^{-1}\).
Testing Sessions One and Two (T₁ and T₂)

Pretesting guidelines were repeated for all remaining testing sessions. During testing sessions one and two (T₁ and T₂), the 42mm Apple Watch (Apple, Inc., Cupertino, CA) was affixed to the participant’s left wrist and the Microsoft Band (Microsoft, Inc., Redmond, WA) was affixed to the right wrist. Once equipped (see section “Equipment Sizing and Preparation”) and calibrated (see section “Equipment Calibration”), a final check was conducted to ensure that all systems, including both smartphones and their paired smartwatches, were synchronized and populating data. For the 42mm Apple Watch, data were recorded using the Apple Health (iOS 9.2.1) platform, and for the Microsoft Band, the Microsoft Band SDK for Android was utilized to record raw sensor outputs. Once all systems were confirmed operational, a new exercise activity was initiated from both smartwatch touchscreens, and a new submaximal exercise test was initiated within the ParvoMedics TrueOne 2400 Metabolic System software (OUSW 4.3.4). Testing session one (T₁) utilized 50% VO₂ max as the target steady-state intensity while session two utilized 75%.

The speed of the treadmill was the primary method to adjust intensity and reach the target VO₂. If adjusting only the participant’s speed yielded an awkward pace (not quite walking nor jogging), adjustments were made to the grade of the treadmill. Fine-tuning of the speed and/or grade were used in order to get and keep the participant as close to the desired steady-state intensity as possible. The transitional period from the time of initiating movement to reaching the target intensity (% of VO₂ max) was estimated at no longer than five minutes, yielding 15 minutes of steady-state intensity. If after five minutes the participant had not reached the target intensity, additional time (as necessary) was added to the session in order to provide at least 15 minutes of exercise data at the desired steady-state intensity.
After 20 minutes from initiation, or 15 minutes after achieving steady state, the test was terminated. The participant’s speed was returned to 2.5 mph for approximately two minutes to allow the participant to cool down. If after the two minutes the participant’s HR remained elevated above 150 bpm, additional time was allotted until it fell below the threshold. Posttesting data collection procedures were then instituted (see section “Posttesting Procedures”).

Testing Session Three (T3)

In testing session three (T3), smartwatch placement was alternated to the opposite wrist. The 42mm Apple Watch (Apple, Inc., Cupertino, CA) was worn on the right wrist (previously left) and the Microsoft Band (Microsoft, Inc., Redmond, WA) was worn on the left (previously right). All other protocols were identical to that of testing session one, at 75% VO2\text{\textsubscript{max}}. The purpose of this third testing session was to compare the data to testing session two (same intensity, 75% VO2\text{\textsubscript{max}}) to assess reliability when the smartwatch was alternated across wrists.

Posttesting Procedures

At the conclusion of each testing session, the treadmill was brought to a complete stop. All equipment was removed and sanitized according to standard laboratory procedures. At this point, the participant was free to leave. The principal investigator then exported EE data from the ParvoMedics 2400 TrueOne Metabolic System software (OUSW 4.3.4), HR data from the Polar H7 HRM, and HR and EE data from the Apple Health (iOS 9.2.1) and Microsoft Health (1.3.20329.1) smartphone applications. All data sheets were exported in the form of a .csv file and securely stored for each testing condition under the participant’s uID and a testing session
identifier. A factory reset was then performed on both smartwatches and their respective smartphone applications. Because the 42mm Apple Watch (watchOS 2.1) took the greatest amount of time to perform this reset (for future sessions), this 42mm Apple Watch always underwent posttesting procedures first. Applicable devices were then placed on their respective chargers for the next testing session, and any remaining testing data on the devices were erased.

Data Analysis

In addition to recording participant and descriptive statistics, a number of statistical tests were performed. Inferential statistics were computed using IBM SPSS Statistics (version 24). Smartwatch heart rate (HR) and energy expenditure (EE) values were analyzed to determine validity (accuracy) as compared to criterion values. Reliability was assessed to determine if devices could be worn interchangeably across left and right wrists. The Polar H7 HRM served as the criterion method for HR and the ParvoMedics 2400 TrueOne Metabolic System served as the criterion method for EE. All tests were run independently for both HR and EE. A 3 (device) x 2 (condition) repeated-measures analysis of variance (ANOVA) \((\alpha=0.05)\) was used to determine significant differences in mean HR (bpm) and mean EE (kCal) for the 42mm Apple Watch and Microsoft Band. Follow-up paired \(t\) tests were run after significant main effects occurred. Pearson correlation coefficients \((r)\) and standard error of estimates (SEE) were then calculated to determine the validity of the HR and EE measured by each smartwatch. Reliability was assessed through intraclass correlation coefficients (ICC) based on a two-way random effects analysis of variance model and an absolute agreement definition \([\text{ICC}(2,1)]\) to determine if devices could be worn interchangeably between the left and right wrists (Shrout & Fleiss, 1979).
RESULTS

Compliance

Of the 26 participants registered to participate in this study, all 26 participants (100%) satisfied the inclusion criteria, were free from all exclusion criteria, and completed the initial VO$_2$$_{\text{max}}$ assessment. Twenty-three (88.5%) participants completed all testing requirements, which allowed them to be included in data analyses. Three (11.5%) participants failed to fully complete all testing requirements, which excluded them from subsequent data analyses. Of these three, one (3.8%) participant terminated participation due to an acute lower extremity musculoskeletal injury not resulting from their participation in the study. The remaining two participants (7.7%) were noncompliant by failing to attend their prescheduled testing sessions for unknown reasons.

Participant Characteristics

Participants consisted of males and females aged from 19 to 37 years old. Participant age represents their age taken during the final testing sessions (Table 1). Height and body mass ranged from 160 to 195 cm and from 50.6 to 98.8 kg, respectively. The modified Balke treadmill protocol yielded VO$_2$$_{\text{max}}$ values between 29.3 and 61.6 mL·kg$^{-1}$·min$^{-1}$; these VO$_2$$_{\text{max}}$ values correspond to ACSM’s percentile rankings (for aerobic fitness) between 3.2 and 98.4%, respectively. Body mass index (BMI, kg·(m$^2$)$^{-1}$), body composition (% body fat), and cellular water content (extracellular water content: total body water) were all assessed through the InBody520 BIA. BMI and body composition varied over notably large ranges: BMI from 17.7 to 32.1 kg·(m$^2$)$^{-1}$ and body
composition between 6.3 and 36.3% body fat. Cellular fluid ratios (ECW:TBW) were determined to be between 0.355 and 0.386, breaking through the lower limit of InBody’s recommended healthy range of 0.360 – 0.390. Mean participant characteristics are reported in Table 1, and a breakout of all participant statistics, including anthropometrics for each testing session, can be found in Appendix J.

Table 1
Participant Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Height</th>
<th>VO₂ max</th>
<th>ACSM Ranking</th>
<th>Body* Mass</th>
<th>Body* BMI*</th>
<th>Body* Fat</th>
<th>ECW:* TBW*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>22.7</td>
<td>174.5</td>
<td>46.6</td>
<td>68.9</td>
<td>75.9</td>
<td>24.8</td>
<td>17.3</td>
<td>0.368</td>
</tr>
<tr>
<td>Median</td>
<td>22.0</td>
<td>175.0</td>
<td>49.2</td>
<td>80.0</td>
<td>76.3</td>
<td>25.6</td>
<td>14.4</td>
<td>0.367</td>
</tr>
<tr>
<td>Std Dev</td>
<td>3.8</td>
<td>9.4</td>
<td>7.6</td>
<td>24.7</td>
<td>13.9</td>
<td>3.7</td>
<td>7.9</td>
<td>0.007</td>
</tr>
<tr>
<td>Min</td>
<td>19.0</td>
<td>160.0</td>
<td>29.3</td>
<td>3.2</td>
<td>50.6</td>
<td>17.7</td>
<td>6.3</td>
<td>0.355</td>
</tr>
<tr>
<td>Max</td>
<td>37.0</td>
<td>195.0</td>
<td>61.6</td>
<td>98.4</td>
<td>98.8</td>
<td>32.1</td>
<td>36.3</td>
<td>0.386</td>
</tr>
</tbody>
</table>

N=23 (16 males, 7 females)
*Mean measurement from all testing trials (3)
1 Body Mass Index
2 Extracellular Water: Total Body Water

Heart Rate (HR)

A 3 (device) x 2 (condition) repeated-measures ANOVA (α=0.05) assessed differences in mean HR (bpm) values (Table 2). Compared to the Polar H7 HRM criterion, mean HR for the 42mm Apple Watch was 0.07% higher for the 50% and 0.12% higher for the 75% VO₂ max conditions, respectively. Mean HR for the Microsoft Band was 2.79% lower for the 50% and 4.26% lower for the 75% VO₂ max conditions, respectively.
Table 2
Mean Heart Rate at 50% and 75% VO₂ max

<table>
<thead>
<tr>
<th>Intensity (Trial)</th>
<th>Polar H7 HRM**</th>
<th>42mm Apple Watch</th>
<th>Microsoft Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% VO₂ max (T₁)</td>
<td>145.1 ± 8.4</td>
<td>145.2 ± 8.4</td>
<td>141.1 ± 8.3</td>
</tr>
<tr>
<td>75% VO₂ max (T₂)</td>
<td>172.3 ± 9.5</td>
<td>172.5 ± 9.3</td>
<td>165.1 ± 8.6</td>
</tr>
</tbody>
</table>

*p = .164; criterion comparison  *p < .0005; criterion comparison
(Follow-up tests are collapsed across intensity)

N=23; α=0.05
**Criterion method for HR
* Significantly different than criterion

The repeated-measures ANOVA (α=0.05) for HR revealed two significant main effects: a device main effect \([F(1.052, 23.141) = 72.3, p < .0005]\) and an intensity main effect \([F(1.0, 22.0) = 111.612, p < .0005]\). The post hoc tests revealed that the Microsoft Band HR was significantly lower than the Polar H7 criterion by an average of 5.6 bpm \((p < .0005)\). In contrast, the 42mm Apple Watch was not significantly different from the Polar H7 criterion \((p = .164)\), with a mean difference of only 0.2 bpm (higher).

Intensity as a main effect was significantly higher by an average of 26.2 bpm \((p < .0005)\) across the increasing intensities. A significant interaction of device by intensity \([F(1.088, 23.930) = 73.318, p = .006]\) was also observed in HR. Figure 2 illustrates the interaction of device by intensity for mean HR measure; difference in mean HR for the Microsoft Band increased from 4.0 (lower) to 7.2 bpm (lower) than criterion values as intensity increased from 50% to 75% VO₂ max.
Figure 2: Device by Intensity Interaction for Mean Heart Rate

HR values for each device were standardized to one reading every five seconds during the exercise protocol / testing trials. The percentage of 42mm Apple Watch HR readings matching the Polar H7 criterion was subsequently determined to be 35.5% (for 50% VO₂ max) and 23.9% (for 75% VO₂ max). Additionally, 88.0% of all Apple Watch HR readings were within ± 10 bpm of the criterion measures. The percentage of Microsoft Band HR readings matching the Polar H7
criterion was 21.9% (for 50% VO\textsubscript{2 max}) and 15.68% (for 75% VO\textsubscript{2 max}); 75.9% of all Microsoft Band HR readings were within ± 10 bpm of the criterion measures.

Energy Expenditure (EE)

A second 3 (device) x 2 (condition) repeated-measures ANOVA (\(\alpha=0.05\)) assessed differences in mean EE (kCals). Compared to the ParvoMedics TrueOne 2400 criterion, mean EE for the 42mm Apple Watch was approximately 2.53% (for 50% VO\textsubscript{2 max}) and 4.12% (for 75% VO\textsubscript{2 max}) lower than corresponding criterion measures. Mean EE for the Microsoft Band was 10.67% (for 50% VO\textsubscript{2 max}) and 17.84% (for 75% VO\textsubscript{2 max}) higher than corresponding criterion values (Table 3).

The repeated-measures ANOVA (\(\alpha=0.05\)) for EE revealed two significant main effects: a device main effect \([F(1.072, 23.583) = 7.715, p = .009]\) and an intensity main effect \([F(1.0, 22.0) = 211.829, p < .0005]\). The post hoc tests revealed that the Microsoft Band EE was significantly higher than the ParvoMedics 2400 TrueOne Metabolic System criterion and 42mm Apple Watch by an average of 22.0 kCals \((p = .028)\) and 28.4 kCals \((p = .005)\), respectively. The 42mm Apple Watch was significantly different (lower) than the ParvoMedics 2400 TrueOne Metabolic System criterion with a mean difference of 6.5 kCals lower \((p = .002)\). Intensity as a main effect was also significantly different (higher) by an average of 57.6 kCals \((p < .0005)\) with increasing intensity. There was no significant device by intensity interaction observed for EE \([F(1.225, 26.959) = 636.880, p = .243]\).
Table 3
Mean Energy Expenditure at 50% and 75% VO$_2$ max

<table>
<thead>
<tr>
<th>Intensity (Trial)</th>
<th>ParvoMedics 2400**</th>
<th>42mm Apple Watch</th>
<th>Microsoft Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% VO$_2$ max ($T_1$)</td>
<td>164.2 ± 51.2</td>
<td>160.1 ± 51.1</td>
<td>182.7 ± 55.9</td>
</tr>
<tr>
<td>75% VO$_2$ max ($T_2$)</td>
<td>221.1 ± 62.0</td>
<td>212.2 ± 55.9</td>
<td>246.4 ± 55.0</td>
</tr>
</tbody>
</table>

*N=23; α=0.05
** Criterion method for EE
*  Significantly different than criterion

Measures of Validity and Reliability

To assess biometric validity, Pearson correlation coefficients ($r$) were calculated for the 42mm Apple Watch and Microsoft Band with 95% confidence intervals (CI) for HR and EE at each intensity (Table 4). Standard error of estimates (SEEs) were also assessed for each biometric (Table 4). HR correlation coefficients for the 42mm Apple Watch ranged between 0.99 and 1.00, with SEE ranging between 0.54 to 1.04 bpm across increasing intensities. Correlation coefficients for the Microsoft Band were considerably weaker, between 0.90 and 0.91, with the corresponding SEE considerably higher, from 3.57 to 4.83 bpm, across increasing intensities. For EE, correlations were lower for both devices. Correlation coefficients for the Apple Watch were measured at 0.98 (regardless of intensity), with an SEE ranging between 10.07 and 12.34 kCals across increasing intensities. The Microsoft Band demonstrated a much weaker correlation at both intensities, between 0.49 and 0.74. SEE was also larger, between 37.60 (for 75% VO$_2$ max) and 45.59 (for 50% VO$_2$ max) kCals, than values for the Apple Watch. In addition to the Pearson correlation coefficients ($r$) listed in Table 4, the corresponding concordance correlation coefficient ($r_c$) of each device is listed in Appendix K (Table 9).
Table 4

Pearson Correlation Coefficients (r) with 95% Confidence Intervals (CI) and Standard Error of Estimates (SEE) for Heart Rate and Energy Expenditure at 50% and 75% VO$_{2\text{max}}$

<table>
<thead>
<tr>
<th>Intensity (Trial)</th>
<th>Heart Rate (HR)</th>
<th>Energy Expenditure (EE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Polar H7 HRM vs. 42mm Apple Watch</td>
<td>Polar H7 HRM vs. Microsoft Band</td>
</tr>
<tr>
<td>50% VO$_{2\text{max}}$ ($T_1$)</td>
<td>$r$ 1.00 (1.00 – 1.00) SEE 0.54</td>
<td>$r$ 0.99 (0.99 – 1.00) SEE 1.04</td>
</tr>
<tr>
<td>75% VO$_{2\text{max}}$ ($T_2$)</td>
<td>$r$ 0.91 (0.80 – 0.96) SEE 3.57</td>
<td>$r$ 0.90 (0.71 – 0.94) SEE 4.83</td>
</tr>
</tbody>
</table>

N=23 (16 males, 7 females)

Reliability agreement was assessed through intraclass correlation coefficients (ICC) based on a two-way random effects analysis of variance model. For HR, device means between the two 75% VO$_{2\text{max}}$ trials (trials two & three) did not differ significantly ($p > .05$), and ICC values ranged from 0.51 to 0.63. The Microsoft Band (ICC = 0.51, 95% CI [0.13, 0.76]) had the weakest reliability across wrists while the 42mm Apple Watch (ICC = 0.61, 95% CI [0.28, 0.81]) and Polar H7 HRM (ICC = 0.63, 95% CI [0.31, 0.82]) had a stronger ICC reliability agreement (Table 5). For EE, device means between the two 75% VO$_{2\text{max}}$ trials did not differ significantly for the Apple Watch ($p = .88$) or Microsoft Band ($p = .80$) but did differ significantly for the ParvoMedics TrueOne 2400 Metabolic Cart ($p = .007$). Agreement in ICC reliability for EE was strongest in ParvoMedics TrueOne 2400 Metabolic Cart (ICC = 0.99, 95% CI [0.98, 0.99]), followed by similarly strong reliability agreement for the Apple Watch (ICC = 0.98, 95% CI [0.96, 0.99])
Reliability agreement was lowest in the Microsoft Band (ICC = 0.73, 95% CI [0.46, 0.88]) (Table 6).

Table 5
Heart Rate Intraclass Correlation Coefficients (ICC) for 75% VO$_2$ max

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Watch (42mm)</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>172.5</td>
<td>9.3</td>
<td>0.61</td>
<td>0.28</td>
</tr>
<tr>
<td>Apple Watch (42mm)</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>174.8</td>
<td>9.1</td>
<td>0.61</td>
<td>0.28</td>
</tr>
</tbody>
</table>

$p = .18$; means do not differ

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Band</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>165.1</td>
<td>8.6</td>
<td>0.51</td>
<td>0.13</td>
</tr>
<tr>
<td>Microsoft Band</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>165.8</td>
<td>10.7</td>
<td>0.51</td>
<td>0.13</td>
</tr>
</tbody>
</table>

$p = .75$; means do not differ

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar H7 HRM</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>172.3</td>
<td>9.6</td>
<td>0.63</td>
<td>0.31</td>
</tr>
<tr>
<td>Polar H7 HRM</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>174.8</td>
<td>9.3</td>
<td>0.63</td>
<td>0.31</td>
</tr>
</tbody>
</table>

$p = .15$; means do not differ

Table 6
Energy Expenditure Intraclass Correlation Coefficients (ICC) for 75% VO$_2$ max

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Watch (42mm)</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>212.2</td>
<td>55.9</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>Apple Watch (42mm)</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>211.9</td>
<td>53.0</td>
<td>0.98</td>
<td>0.96</td>
</tr>
</tbody>
</table>

$p = .88$; means do not differ

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Band</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>236.4</td>
<td>55.0</td>
<td>0.73</td>
<td>0.46</td>
</tr>
<tr>
<td>Microsoft Band</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>244.2</td>
<td>57.1</td>
<td>0.73</td>
<td>0.46</td>
</tr>
</tbody>
</table>

$p = .80$; means do not differ

<table>
<thead>
<tr>
<th>Device</th>
<th>Intensity (Trial)</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParvoMed 2400</td>
<td>75% VO$_2$ max (T$_2$)</td>
<td>221.1</td>
<td>62.0</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>ParvoMed 2400</td>
<td>75% VO$_2$ max (T$_3$)</td>
<td>224.5</td>
<td>63.4</td>
<td>0.99</td>
<td>0.98</td>
</tr>
</tbody>
</table>

*p = .007; means differ

* Means are significantly different between trials
DISCUSSION

The purpose of this study was to assess the biometric validity and reliability (across left and right wrists) of heart rate (HR) and energy expenditure (EE) as recorded by the first-generation Apple Watch (42mm) and Microsoft Band during submaximal aerobic exercise at steady state. Validity was assessed through Pearson correlation coefficients \( (r) \) and standard error of estimates (SEE) for two testing conditions at increasing aerobic workloads (from 50% to 75% \( \text{VO}_2 \text{max} \)). Workload was increased to determine if intensity had a significant influence on biometric accuracy output. A third testing session (at 75% \( \text{VO}_2 \text{max} \)) was performed to assess agreement in device reliability across left and right wrists (and vice versa). This type of reliability was assessed through the strength of intraclass correlation coefficients (ICC).

The findings of the current study demonstrate that the 42mm Apple Watch is a valid smartwatch for the measurement of HR at 50% and 75% \( \text{VO}_2 \text{max} \) during submaximal aerobic treadmill work at steady state. A correlation coefficient of \( \geq 0.90 \) and SEE at \( \leq 5 \) bpm has been used as a standard model for the assessment of valid measures of HR in various research studies (Leger & Thivierge, 1988; Terbizan, Dolezal, & Albano, 2002). Other studies have determined validity for HR by using larger ranges of SEE at \( \leq 10 \) bpm (Lee & Gorelick, 2011). Furthermore, the Pearson correlation coefficient of 1.00 (for the Apple Watch in the present study) represents the highest attainable correlation strength, as well as one of the highest correlation strengths observed in any device that records HR; this demonstrates exceptional HR validity for the Apple Watch. The Microsoft Band was not as valid as the Apple Watch, though correlation coefficients
were still observed within an acceptable range for 50% \( \text{VO}_2 \text{max} \). Microsoft Band validity was questionable at 75% \( \text{VO}_2 \text{max} \) when considering the lower limits of the 95% CI.

The findings of the Apple Watch providing valid measures of HR are consistent with recent research conducted by Wang et al. (2016), who assessed the accuracy of four different wrist-worn heart rate monitors. As compared to an ECG/EKG criterion at five different treadmill speeds, the Apple Watch had the highest concordance correlation coefficient \( (r_c) \) of any wrist-worn heart rate monitor (Wang et al., 2016). Both the aforementioned research study and the present investigation reported Pearson correlation coefficients (present study only) and concordance correlation coefficients (both studies) for the Apple Watch at \( \geq 0.91 \) (Wang et al., 2016). The referenced study has demonstrated superior HR accuracy during rest, which this study did not assess (Wang et al., 2016). The present study agrees with the findings of Wang et al. (2016) that the Apple Watch provides highly accurate HR measurements and, to best of the investigator’s knowledge, is the first research study to assess the continuous measurement of HR in both the Apple Watch and wrist-worn optical HR monitors in general.

Overall, the Apple Watch measured HR with exceptional accuracy, considerably higher than that of the Microsoft Band during aerobic work. Even still, SEE, or typical error, remained within \( \pm 5 \) bpm for both devices at any given intensity. These findings using the above-mentioned criteria are on par with many other HR validation studies (Lee & Gorelick, 2011; Leger & Thivierge, 1988; Terbizan et al., 2002). While the validity of optical heart rate sensors across different current-generation wearables has been inconsistent at best (Spierer, Rosen, Litman & Fujii, 2015; Valenti & Wsterterp, 2013), the current findings add to a growing body of evidence that use of the Apple Watch provides valid measures of optical heart rate tracking at different
speeds and intensities (Wang et al., 2016). While HR correlation strength for both devices declined as a function of increasing exercise intensity, a decrease in this regard is both typical and to be expected with the use of any HRMs (Lee & Gorelick, 2011; Leger & Thivierge, 1988). It is noteworthy, however, that the observed drop in correlation strength for the Apple Watch was miniscule (almost nonexistent), which is quite remarkable.

Overall, Apple Watch HR measurements (on average) were not observed to differ from the criterion values, whereas Microsoft Band HR measurements were observed to be significantly lower. The percentage of Apple Watch HR readings \( n = 4140 \) found to be within \( \pm 10 \) bpm of criterion values was greater in the Apple Watch (88.0\%) than in the Microsoft Band (75.9\%). Additionally, the percentage of exact criterion matches for the Apple Watch was between 23.9\% and 33.5\% and between 15.7\% and 21.9\% for the Microsoft Band. The percentage of exact matches declined in both devices as a function of increasing intensity. These findings further support the practice of using the Apple Watch for the valid measurement of HR.

In the case of reliability, intraclass correlation coefficients (ICCs) measured the strength in a device’s ability to be switched from the left to right wrists (or vice versa) while maintaining consistent HR readings. ICC agreement interpretations were based on these criteria (ranges): ICC > 0.75 indicates a very good to excellent reliability agreement, 0.4 < ICC < 0.75 indicates a fair to good reliability agreement, and an ICC < 0.4 indicates a poor reliability agreement (Portney & Watkins, 2000). Additionally, the strength of reliability agreement for each device was expressed using a range corresponding to that of the lower and upper limits of the 95\% CI. The Apple Watch had very similar reliability agreement to that of the Polar H7 criterion. It also had a reliability agreement (ICC = 0.61, 95\% CI[0.28, 0.81]), ranging from poor to very good, while the Microsoft
Band had a weaker agreement (ICC = 0.51, 95% CI[0.13, 0.76]), also classifying the band as poor to very good. While none of the mean HR values in T2 differed significantly from those of T3, ICC agreement had insufficient strength across the 95% CI to confidently classify any wearable device as reliable across wrists. It should also be noted that while the reliability assessment in this study was intended to measure biometric agreement across wrists, the ICC reliability coefficients reflect measurement error due to wrist location of watches and day-to-day variability. It was not possible to separate day-to-day error from location error.

As with HR, the 42mm Apple Watch outperformed the Microsoft Band in the valid measurement of EE at both exercise intensities. Pearson correlation coefficients ($r$) for the Apple Watch remained closely correlated to the criterion values within a 95% CI for each intensity level. These findings, along with the high concordance correlation coefficients in Table 9 (Appendix K) are quite remarkable and help to establish the Apple Watch as an exceptionally valid measure of EE across intensities. On average, the Apple Watch’s tendency to underestimate EE at approximately 96 to 98% of criterion values was insignificant. SEE, or typical error, were between 10.07 kCals (50% VO$_{2 \text{max}}$) and 12.34 kCals (75% VO$_{2 \text{max}}$), and it is worth noting that most consumers would likely prefer EE to be underestimated rather than overestimated.

The Microsoft Band was considerably less valid in EE measurement at both intensities than the Apple Watch. Mean EE values for the Microsoft Band were determined to be approximately 111% of the criterion values for both of the workloads/intensities, and Pearson correlation coefficients, along with concordance correlation coefficients (Table 9 in Appendix K), were notably weaker at 50% VO$_{2 \text{max}}$ than for the higher 75% VO$_{2 \text{max}}$ condition. The small lower limits of the 95% CI for the correlation coefficient also caused concern. Because an atypical drop in SEE
from 45.59 kCals (at 50% VO$_{2\text{max}}$) to 37.60 kCals (at 75% VO$_{2\text{max}}$) was observed, this may also raise questions about its ability to provide valid EE measurements.

Device reliability agreement for EE was assessed through the use of ICCs and 95% CIs. In contrast to the HR ICC, strength of EE ICCs were found to be markedly stronger. Overall, the Microsoft Band reliability agreement (ICC = 0.73, 95% CI[0.46, 0.88]) ranged between poor and very good but was much weaker than the agreement of the Apple Watch (ICC = 0.98; 95% CI[0.96, 0.99], which was excellent. While the reliability agreement of the Microsoft Band ranged from poor to very good, its validity evidence for EE was not good. From this assessment, the Apple Watch is considered to be the more exceptional device based on validity and reliability evidence for EE.

While these findings are limited to the context of this research study (aerobic submaximal exercise), they still adequately demonstrate the significance of smartwatch (or wearable technology, in general) validation as an important area of research. The findings from the current study demonstrate how two seemingly similar devices can produce very different results in practice. While the Microsoft Band did a good job at measuring HR and a poor to fair job at measuring EE, the Apple Watch was remarkable in its ability to accurately measure both HR and EE. Moreover, while the Apple Watch had insufficient reliability agreement across left and right wrists for HR, it had exceptional agreement for EE. Additionally, with the Microsoft Band consistently overestimating HR in the 75% conditions, a lower mean HR value would theoretically produce a lower caloric expenditure (or EE). Nonetheless, the opposite was observed. This may help demonstrate how device engineering and algorithms are critical to the overall performance of the devices.
Limitations

There were several challenges and limitations associated with this type of research study that may not be immediately apparent to readers and prospective researchers. First, the decision to use the Polar H7 HRM, an electrode-based chest strap, in place of a typical EKG criterion may not be supported by all researchers. The Polar H7 HRM, specifically, has been substantiated by Wang et al. (2016), who found a concordance correlation of 0.99 (Polar H7 HRM with the EKG criterion). Other research has supported similarly high findings of validity in various other Polar HRMs for HR assessment. Although measurements of HR validity among most devices tend to decline with an increase in speed, high validity has been established at rest and during aerobic exercise when using electrode-based Polar HRMs (Lee & Gorelick, 2011; Terbizan et al., 2002).

The development of an experimental design which is capable of assessing independently operating wearable devices (and their biometrics) in parallel can be a very complex task. One reason for this is because some devices are targeted to work with a specific subset of mobile platforms which may consequently hinder capabilities or present unforeseen obstacles. Because different mobile ecosystems are developed with a range of different capabilities, biometric recording and data extraction may be limited to specific mobile/wearable device implementations. Wang et al. (2016) pointed out that continuous measurement of HR (which was assessed in the present study) enables more detailed comparisons to be made. Unfortunately, continuous measurement cannot be done with all wearable devices.

For the present study, this limitation was observed specifically in the Apple Watch. Although it provided superior biometric accuracy and reliability of both HR and EE, it was incompatible with non-iOS-enabled devices; this limited the data recording and extraction protocol
to working within the capabilities of iOS’s native Health Application and presented additional challenges in extracting and storing sensor data. It also provided a major hurdle in terms of heart rate sampling rate. Specifically, the Apple Watch had the slowest sampling rate of all HR devices at once every five seconds. All other devices were capable of measuring HR, at minimum, once per second. Consequently, scaling the sampling rate to the slowest device was an important consideration in maintaining uniform data across devices. For this specific study, the limitation of being forced to work within Apple’s ecosystem resulted in multiple smartphones being required for data collection and introduced a number of hurdles to maintaining consistency in the experimental design. While the Apple Watch can provide useful and accurate measures of HR for a healthy population (from an activity-tracking standpoint), this cannot be generalized for the evaluation of monitoring HR in diseased or older populations. Additionally, while the low sampling rate of the Apple Watch alone should be enough support that the device should not be used to monitor diseases or medical condition, this research was conducted on healthy individuals who were screened for their participation in physical activity.

The amount of raw data associated with this type of experiment should also be noted. Even though HR sampling rate was scaled to once every five seconds, more than 12,000 individual HR values needed to be checked and analyzed. The amount of processing required in creating unified datasheets that could be used in analyses was quite cumbersome. Simply put, there are tons of device-specific variables to consider, and with new iterations of devices emerging each year, there is no handbook that could possibly prepare someone for the challenges of performing wearable research. Though the difficulty associated with conducting such a study will certainly vary
depending on the device being used (as well as *how* it is being used), these hurdles may very well contribute to a lack of research in the area of wearables and biometric data.

Given the increasing number of wearable devices hitting the market and a growing number of biometrics available for analysis, research possibilities with wearables are increasing every day. Given the exceptional validity of both biometrics as reported by the 42mm Apple Watch (a first-generation wearable smartwatch), this device may pique the interest of researchers looking to add to the findings of this study, or biometric validity testing in general. With the Apple Watch Series 2 being released in September of 2016, it may also be a promising target for future biometric research. Perhaps it could even shatter some of the ecosystem challenges/limitations (such as HR sampling rate) that were encountered in this investigation. For researchers interested in further exploring the Microsoft Band, a second-generation Microsoft Band was released by the company in October of 2015, boasting a more-advanced specification set. Unfortunately, Microsoft has recently pulled these devices from their online store, and future iterations of the device seem questionable.

**Conclusion**

The present study discovered that Apple’s first-generation wearable device shows tremendous promise as a valid activity/exercise tracking device (for recording energy expenditure and heart rate). The device is both well-built and exceptionally accurate at tracking heart rate and energy expenditure during aerobic exercise at different intensities. It is also exceptionally reliable at tracking energy expenditure, but not as reliable as heart rate, when the device is worn on different wrists on different days. Conversely, while the Microsoft Band was moderately accurate in heart rate measurement, it was neither accurate nor reliable in the assessment of energy
expenditure. The HR reliability findings suggest that wearable smartwatches should always be worn on the same wrist in order to obtain the most valid measures.

With the added convenience and portability associated with these devices, the findings from this study are certainly welcome to those looking to expand their activity tracking or monitor their biometrics. Even in spite of potential hurdles associated with wearable research, the wearable technology sector as a whole holds huge promise for observing and monitoring a number of useful biometric data. With more research that aims to establish validity/accuracy (as was the intention of this study), consumers can become more educated about specific wearable devices, and eventually older / conventional exercise / activity-tracking equipment may be phased out entirely.
REFERENCES


APPENDIX A

REVIEW OF LITERATURE
REVIEW OF LITERATURE

Biometric Data

Much of the world, including the Federal Drug Administration (FDA) and some businesses, agree that the use of wearables is both convenient and beneficial to general wellness promotion ("FDA Device Guidance: General Wellness: Policy for Low Risk Devices - Policy and Medicine," 2015). Health insurance companies are going as far to offer reduced insurance premiums in exchange for access to these data, citing the notion that smartwatches and activity trackers may help promote a greater level of physical fitness and promote stress management ("FDA Device Guidance: General Wellness: Policy for Low Risk Devices - Policy and Medicine," 2015). The FDA released a Draft Guidance of the current policy for these devices in 2015, confirming this viewpoint ("FDA Device Guidance: General Wellness: Policy for Low Risk Devices - Policy and Medicine," 2015). What is particularly troublesome is that the same document suggests a current lack of federal regulation for these devices. Further, the FDA intends to continue its oversight of wearable device regulation as long as manufacturers avoid making claims that their devices are able to help diagnose/treat obesity (or any other disease) ("FDA Device Guidance: General Wellness: Policy for Low Risk Devices - Policy and Medicine," 2015). While this reduces the headache for manufacturers to market their products, it also frees them from any liability of producing reputable (biometrically accurate) wearables. With an overall lack of scientific validation for current wearable biometrics, it now becomes the job of the manufacturer or independent researchers to (optionally) validate these devices.
Although each wearable has a unique array of sensors, two sensors provide basic feedback about a user’s movement: the accelerometer and gyroscope (Gleadhill, Lee, & James, 2016). The accelerometer provides information about linear acceleration while the gyroscope measures rotational/angular velocity (Dobkin & Dorsch, 2011). This sensor combo is ultimately responsible for determining if a user is in motion, the type of motion, and how fast it is occurring (Dobkin & Dorsch, 2011). These sensors fuel the algorithms that tell the device if the user is moving or standing still while also assessing speed and distance traveled (Gleadhill et al., 2016). When not in motion, these two sensors can differentiate between seated and lying positions and are frequently called upon to analyze small movement patterns for estimating a user’s quality of sleep (Bryson, 2009).

Since these two sensors are solely responsible for measuring movement, they are core elements of any wearable device with biometric capabilities (Gleadhill et al., 2016). Barring the development of any new revolutionary sensor, they will continue to be an integral part of every tracking device (Bryson, 2009). Lately, newer smartwatches have also begun incorporating additional sensors to refine metrics ("Wearables: Fad or the Future?" 2015). The inclusion of a GPS tracker, for example, can be used to provide the pin-point accuracy needed to correct traditional errors of linear acceleration and distance. Some wearables even showcase a built-in barometer to precisely measure changes in elevation and atmospheric pressure. Most recently, however, almost every premium smartwatch manufacturer has incorporated an optical heart rate (HR) sensors in their device (Tamura Maeda, Sekine, & Yoshida, 2014). These sensors measure HR by analyzing small changes in blood flow and capillary perfusion at the wrist (Tamura et al., 2014). In order to calculate any of these metrics, it is also necessary to input accurate
anthropometric data. The algorithms compute both basic movement statistics and more advanced physiological parameters (such as HR and EE); such metrics are computed through algorithms which are dependent on the input of various parameters and constant biosensor feedback (Tamura et al., 2014).

Traditionally, these biometric data were obtained manually and the continuous collection of many metrics was either tedious or flat-out impractical. Measurement of HR was accomplished through the palpation of arteries and SC was collected through simple human observation (Laporte, Montoye, & Caspersen, 1985). While the measurement of more advanced metrics (such as EE) has been possible since the 1860s, until just a few decades ago, expensive urine/saliva tests or whole-room calorimeters were necessary to obtain these values (McLean & Tobin, 1987). Today, a number of different devices have made the continuous measurement of such parameters much more practical (Hills, Mokhtar, & Byrne, 2014).

Thanks to modern technological advancements and years of testing, the precise and continuous measurement of important physiological variables (such as HR and EE) is more possible than ever (Hills et al., 2014). Many of these technologies have also been validated over several decades and numerous research studies (Hills et al., 2014). While many methods have been assessed and proven more than capable of delivering, the most precise technologies often come with a premium price tag, usually outside the reach of the average consumer (Hills et al., 2014). When cost is not a barrier, modern equipment is often bulky or obtrusive, especially when used during physical activity. Therefore, the use of such equipment is not necessarily conducive to everyday use. While wearable technology has certainly enhanced the accessibility of these metrics,
it is simply not yet known if, and to what degree of accuracy, these wearables can measure biometrics such as HR and EE.

Heart Rate (HR)

Literally vital to life, heart rate (HR) is one of the most basic, yet critical biometric markers to consider when assessing human health and physiology (Klabunde, 2012). Though HR is simply the number of times the heart contracts and relaxes per minute, the implications of a failing heart often manifest themselves through tachycardia (excessively rapid heart rate); ignoring warning signs for HR could quite possibly be the difference between life or death (Farrell, Joyner, & Caiozzo, 2012). A lower HR during both rest and aerobic exercise is a simple indicator of greater cardiovascular health and decreased risk of coronary artery disease (CAD; Farrell et al., 2012). CAD, or the buildup of plaque on the inner vascular walls, is also the most common form of heart disease (Klabunde, 2012). Further, according to the National Heart, Lung, and Blood Institute, CAD is the leading cause of death in the US (Klabunde, 2012).

Of all the metrics smartwatches can currently assess, HR is one of the most (if not the most) vital biometric due to the critical health implications. Physiologically, HR is also a great (and one of the most accurate) determinants of physical exertion (Farrell et al., 2012). In nearly any population, the monitoring of HR is something that can provide invaluable information. Since most wearables are now equipped with the ability to continuously monitor HR at the wrist, the accuracy of monitoring HR through this technique must be substantiated.
Electrocardiogram (ECG)

Electrocardiograms (ECGs) are one of the many tools available for monitoring HR but more commonly used to monitor cardiac function during medical (surgical) procedures or diagnostic screenings for cardiovascular abnormalities (Klabunde, 2012). By plotting voltage amplitude (y-axis) over time (x-axis), cardiologists are able to analyze changes in electrical flow (in as little as 40ms intervals) to detect pathologies in a number of different waveforms throughout the cardiac cycle (Burch & DePasquale, 1990; Klabunde, 2012). While the nature of an ECG is typically diagnostic, they also provide the most accurate measurement of rhythm by accounting for small fluctuations from beat to beat (Malik et al., 1996). ECGs are the most sensitive tool for detecting both HR as well as changes in HR over time (Malik et al., 1996). The analysis of small spikes/drops in HR are useful when analyzing stress-response patterns, including those occurring during cardiovascular stress (Malik et al., 1996).

Twelve-lead ECGs are the gold standard for cardiovascular diagnostics and provide continuous and precise monitoring of different myocardial structures (Klabunde, 2012). Variations in lead configurations are implemented to conduct cardiovascular testing for different purposes, and the flexibility in these configurations allows for selective trade-offs between precision and user mobility (Malik et al., 1996). Lead variations are particularly useful when continuous monitoring of the entire myocardium is not necessary but the highest precision in instantaneous HR is desired (Malik et al., 1996). In scenarios where movement is required during testing, different lead configurations allow leads (and their corresponding structures) to be dialed down by selectively emphasizing the most important myocardial structures (Farrell et al., 2012). As the gold standard for both accuracy and precision in electrophysiology, they often provide more than what
is necessary in otherwise healthy individuals (Malik et al., 1996). Monitoring of HR through the use of a 12-lead configuration is not necessary if myocardial rhythm (HR) is the primary goal (Klabunde, 2012).

Heart Rate Variability (HRV)

In part, ECGs are considered the gold standard due to their ability to measure the oscillation in the interval between consecutive heartbeats or cardiac cycles and the oscillation between instantaneous HR (Malik et al., 1996). The variations in oscillations are more commonly referred to as heart rate variability (HRV; Malik et al., 1996). HRV most generally analyzes the variability between beats, or between each R-R interval of the cardiac cycle (Malik et al., 1996). HRV is of particular importance when looking at any underlying pathological conditions (Leger & Thivierge, 1988; Malik et al., 1996). The small beat-to-beat changes occurring between two peak ‘R’ waveforms are critical to providing a highly accurate and real-time analysis of not only HR but variations in HR (Malik et al., 1996). When tracking instantaneous changes in HR, HRV is crucial; however, if average HR over an extended duration is desired, HRV will not play as major of a role (Malik et al., 1996). While ECGs are the primary medium for assessing HRV, some HRM chest straps have been validated as “ECG-Accurate,” meaning they have the ability to detect the small fluctuations in the R-R intervals, just as an ECG would (Laukkanan & Virtanen, 1998). These fluctuations in HRV can be identified in activities such as respiration, where HRV is normal; during inspiration vagal tone is released causing a drop in parasympathetic nervous response and ultimately elevating HR (Klabunde, 2012). During expiration, the opposite holds true, and HR slows as vagal tone is increased (Klabunde, 2012).
Heart Rate Monitors (HRM)

Similar to ECG setups, chest strap heart rate monitors (HRM) are perhaps the most commonly used tool to provide an accurate measurement of HR. These configurations are used by placing an elastic strap containing electrodes around the torso, directly under the manubrium (Gamelin, Berthoin, & Bosquet, 2006). The strap’s electrodes are used to sense impulses like that of a standard ECG, but instead of isolating different leads to diagnose different pathologies, these straps are attached to transmitters which relay information such as the frequency of a repeating R-R interval through an attached module (Gamelin et al., 2006). Unlike ECGs, which listen for electrical activity at specific leads, HRMs do not isolate any particular anatomical structures or otherwise provide measurable voltage changes (Gamelin et al., 2006). Because these devices are used frequently, modules are often coded so the transmission module only communicates with a matching (or coded) receiver.

Polar Electro Oy of Kemple, Finland, has been conducting validation studies on their various chest-strap HRMs since 1984, and they have produced validation data from individuals as young as three years of age demonstrating a high correlation (to that of an ECG) during rest as well as a number of different types (and durations) of physical activity. To mention a few, Polar Electro found a high correlation at rest for children \( r=0.99 \) between the R-R intervals reported by their Polar S810 HRM and a standard ECG for twelve children aged 9.6 ± 0.9 years of age (Gamelin et al., 2006). When a Polar H7 Bluetooth HRM was paired with the Polar V800 wrist-watch as a receiver, the H7 was capable of detecting HR and changes in the R-R interval, at rest, with only a 0.086% error rate (Giles Draper, & Neil, 2015). Polar’s H7 Bluetooth HRM has also
been shown to report HR with a high correlation to that of an ECG ($r = 0.99$) and a pulse oximeter ($r = 0.97$) in supine, seated, and prone positions (Cheatham, Kolber, & Ernst, 2015).

**Pulse Detection**

Another method of assessing HR is through the use of pulse detection mechanisms. This method is less commonly employed when continuous HR measurements are desired because the method requires the user to hold steady to take measurements (Hashem, Shams, Kader, & Sayed, 2010). Pulse-detection watches have been investigated for accuracy at rest and less commonly during exercise (Hashem et al., 2010). These systems work by the user placing their index finger and thumb on ECG leads of the opposite limb (Hashem et al., 2010). The detection system is mostly used to assess heart rate intermittently or following an activity (Hashem et al., 2010). Because these systems require constant points of contact to provide measurement, they are less commonly incorporated into wristwatches where continuous HR measurement is desired. Nonetheless, these systems have shown to be mostly accurate when compared to traditional ECG and HRM measurements, with a tendency to decline in accuracy as speed or exercise intensity rises (Lee & Gorelick, 2011).

In one study which examined the validity of the Smarthealth watch, the pulse watch was used simultaneously with a Polar Vantage XL and an ECG. Pearson correlation coefficients between the three devices were calculated (Lee & Gorelick, 2011). Measurements taken during rest and intermittently during exercise on a treadmill (at speeds between 2.0 mph and 4.0 mph) yielded the correlation ($0.93 < r < 0.97$) and showed a slight reduction when speed was increased to 6.0 mph ($r = 0.81$). Although a drop was observed at higher speeds, the ECG-sensing pads on
the Smarthealth watch were considerably accurate during all five testing conditions (Lee & Gorelick, 2011).

Optical Sensing

Optical HR sensors are the latest in HR tracking technology and measure HR through the use of photoplethysmography (PPG). More simply, these units work by emitting green and (sometimes) infrared light (Parak & Korhonen, 2014). LEDs emit light onto the skin in close proximity, and receptors analyze light displacement (or reflection) after it absorbed by the blood (Parak & Korhonen, 2014). The optical sensing units consist of various components but are ultimately constructed of some type of emitting component and receptor component (Parak & Korhonen, 2014). The amount of light processed by the sensing units can also undergo a number of various algorithm-induced adjustments as necessary (Seunghoon, 2014). In turn, optical sensing units are sensitive to the design of the unit as a whole, including the unique construction of the wearable device that houses them, and individual characteristics.

In low-light situations and during vigorous activity, some optical sensing units receive input to amplify light or receptor activity to compensate for changes that may arise (Seunghoon, 2014). This input can be triggered from a loose-fitting device or any number of obstructions such as sweat or arm hair (Seunghoon, 2014). Ultimately, PPG integrates the emitting and sensing components to measure small changes in capillary density and vascular perfusion, ultimately estimating HR (Seunghoon, 2014). Many units are constructed with an optical emitter utilizing at least two LEDs of green (or infrared) and a digital signal processor (DSP) containing the receptors (Parak & Korhonen, 2014). In addition to this, other device sensors (most notably the
accelerometer) can sometimes feed into the DSP to provide input about user motion. Through algorithms, the DSP then best adjusts to different conditions to provide the most accurate numerical output for HR (Tamura et al., 2014).

Philips manufactures an optical HR sensor that is present in a few consumer devices such as the Mio Alpha Optical HRM and the Schosche Rhythm (Parak & Korhonen, 2014). For these devices in particular, the same Philips sensor was shown to provide a different result depending on the implementation of the unit. The validity of the Philips optical sensor, as compared to an ECG construct, yielded a range of accuracy between 29% and 83% (Parak & Korhonen, 2014). The accuracy of the HR measurement varied both by device and by activity type (Parak & Korhonen, 2014). While these units have the ability to take into account movements of the user, due to their location (usually on the wrist), the optical units are particularly susceptible to “artifacting” when the user sweats and/or produces excessively jerky movements (Spierer, Rosen, Litman, & Fujii, 2015). These systems have shown unpredictability to variations in a user’s skin tone and skin thickness (Valenti & Wsterterp, 2013). Additionally, tattooed/inked skin typically does not play well due to interference of light transmission (Spierer et al., 2015). Due to a number of many different potential confounds when using these systems, they must be thoroughly vetted in a variety of environments under a number of conditions to determine specifically which conditions they are and are not suitable for use in.

Earlier this year, the popular manufacturer Fitbit, Inc., got slapped with a class-action lawsuit for alleged “wild fluctuations” in their optical sensing units (McLellan et al. v. Fitbit, 2016). The plaintiff claimed that Fitbit’s PursePulse technology, which is in fact Fitbit’s optical heart rate sensing system on their recently introduced Charge HR and Surge smartwatches, yielded
sporadic and wildly inaccurate values (McLellan et al. v. Fitbit, 2016). Those who carried out the lawsuit claimed that PurePulse was at times up to 75 bpm off of the actual HR value (McLellan et al. v. Fitbit, 2016). What’s more, these readings showed a notable degradation as the individual exercised or increased exercise intensity. This is a great example of how poorly designed, or poorly implemented optical heart rate sensors can provide inconsistent and inaccurate results.

While it is possible to achieve an acceptable level of accuracy with some of these optical sensing units (during lower intensity activity), the engineering and implementation of these units needs to be sound and tested extensively. Without this testing, it opens the potential for a number of major issues regarding accuracy and consistency. Particularly due to the fact that most current-generation smartwatches and fitness trackers utilize optical sensing systems, these need to be extensively tested on an individual basis.

Energy Expenditure (EE)

As was mentioned earlier, some of the earlier iterations of activity trackers were researched using a large sample of 30 men and 30 women. These individuals underwent a 69-minute exercise protocol wearing eight different activity trackers, and the percent error for the caloric expenditure as reported by the devices was shown to be between 10.1% and 23.5% (Lee et al., 2014). Earlier iterations of fitness-tracking devices still relied on proprietary algorithms to compute caloric output but did not contain the heart rate sensors that today’s smartwatches are equipped with.

While clearly some previous-generation fitness wearables have shown the ability to achieve a considerable amount of accuracy using their algorithms, it should be very clear that the devices’ ability to calculate these values for energy expenditure are purely estimations. These are
the result of the user’s manually logged anthropometric information including height, weight, and age. It goes without saying then, that while these algorithms will likely not yield a 100% accuracy for energy expenditure, any potential accuracy in this department is completely contingent upon the user’s ability to enter proper metrics before allowing the device to record.

As with the wearables research conducted by Lee et al. (2014), any validation will be best served by comparing the reported data against a metabolic system, such as the ParvoMedics Trueone 2400 Metabolic System (ParvoMedics, Salt Lake City, UT), to provide a detailed overview of the individual’s O2 and CO2 content (Farrell et al., 2012). By utilizing this method, the subject can obtain an accurate construct measure for EE (Farrell et al., 2012).
REFERENCES


APPENDIX B

INFORMED CONSENT FORM
I do hereby consent to take part in a research study that assesses the accuracy of two wearable fitness trackers during sub-maximal exercise by researchers at Northern Illinois University. William Kalamaras, a graduate student of Exercise Physiology at Northern Illinois University (NIU) is conducting the study under the direction of Dr. Amanda Salacinski from the Department of Kinesiology and Physical Education. The research study will assess the accuracy of two top-selling wearable smart watches that double as activity trackers – the Apple Watch and Microsoft Band. The primary purpose of the research is to provide a scientific assessment of these device’s capabilities to determine if they could potentially be used to substitute other devices and enhance user accessibility.

As a participant of the study, I understand that I will be required to complete two different medical history questionnaires that will be used to screen me for participation in the study. I understand that my objective will be to run on a treadmill on 5 different occasions for a total time commitment of 3.5 hours. I understand that I will be asked to wear a VO$_2$ mask, heart rate transmitter, Apple Watch, and Microsoft Band simultaneously during exercise in each session performed in the lab. I understand that I will have my height, weight, and body composition analyzed during each testing session as a participant. I understand that the data collected from every and all of these devices will be securely and confidentially stored. I understand that all collection of this data will take place in the Human Performance Lab in 206 Anderson Hall at NIU. I understand that the data obtained from my participation will be used in multiple data analyses, even after testing and data collection has concluded.

I understand that prior to wearing the Polar heart rate monitor, Apple Watch, and Microsoft Band, I will be required to perform a preliminary VO$_2$ max test on the treadmill wearing an oxygen consumption mask and a Polar heart rate monitor. I understand that this test will require me to run on a treadmill with incrementally increasing intensity until I can no longer continue. I understand that this may take approximately 45 minutes to complete, and may be uncomfortable at times. I understand that after the initial assessment, I will need to schedule four additional testing times 36-48 hours apart from one another which will each require me to run on treadmill for 20 minute increments. I understand that during these testing sessions, the incline of the treadmill will remain at a 0% grade, but the speed will be varied by the researcher in order to keep me at the predetermined workload that was determined using the preliminary VO$_2$ max test. I understand that collectively, the 45-minute pretest and three subsequent 20-minute testing sessions (including an additional 15-20 minutes for preparation) will take approximately 3 hours of my time over a total of 5 sessions to complete.
I understand that my participation in this study is completely voluntary and that I may withdraw from this study at any time without penalty or prejudice. I understand that if I have any additional questions concerning this study, I may contact William Kalamaras at (815) 302-7277 (wkalamaras@niu.edu), Amanda Salacinski at (815) 985-7289 (asalacinski@niu.edu), or one of the faculty advisors. I understand that if I wish to obtain further information regarding my rights as a research subject, I may contact the Office of Research Compliance at Northern Illinois University at (815) 753-8588. I understand that all records are held in confidence and that my name will not be used on the final report or associated with any data. Only NIU qualified research personnel listed on this will see the data and/or be present at the time of the study. Any information obtained in connection with this study and that may identify me individually will be kept confidential at all times.

I understand that the risks from participating in this study include mild soreness and fatigue from physical exertion. I understand that this may occur during my VO$_2$ max assessment, or during one of the additional testing sessions. I understand that per Northern Illinois University’s policy, I will not be provided compensation for my involvement, nor does the University carry insurance to cover injury or illness incurred as a result of my participation in University sponsored research. I understand that should I suffer a minor injury, that all subjects will be referred to their Primary Care Physician, NIU Health Services, or the nearest hospital and in the event of serious injury, emergency medical services will be notified immediately.

I understand that the benefits of this study include determining my body composition and my VO$_2$ max, and will provide information that may assist me in making a more educated purchase of a fitness tracker in the future. I also understand that because I will be exposed to these devices first hand, this may give me the opportunity to see how they work. I understand that these tests and their results will be provided to me at no charge, and that my test results will be furnished after the completion of the research study.

I understand that my signature below indicates that I have read this document in its entirety and will act as my consent to participate in the wearable technology validity study. I understand that by signing this form I have had the opportunity to directly inquire about further questions with the principle researcher. I understand that my consent to participate in this project does not constitute a waiver of any legal rights or redress I might have as a result of my participation, and I acknowledge that I have received a copy of this consent form.

(continued on next page)
Consent to be video recorded and photographed (optional)

Wearable Technology Research Study

As an optional supplement to this research study, I wish to opt-in to being video recorded and photographed before, during, and after my participation as a research participant. I do hereby authorize that my signature below will serve as an agreement for me to be digitally recorded using video, sound, and photography at any point throughout my participation as subject in research testing. I authorize that the captured footage or photography may be released for use in a limited capacity for research documentaries and/or research group promotional videos. I understand that should I change my mind after signing this authorization, I can provide a written, signed-request to have this authorization reversed.

Name of Participant (printed): _______________________________  Date: ___/___/2016

Signature of Participant: _______________________________  Date: ___/___/2016

Name of Investigator: William Kalamaras  Date: ___/___/2016

Signature of Investigator: _______________________________  Date: ___/___/2016
APPENDIX C

PAR-Q
PAR-Q & YOU
(A Questionnaire for People Aged 15 to 69)

Regular physical activity is fun and healthy, and increasingly more people are starting to become more active every day. Being more active is very safe for most people. However, some people should check with their doctor before they start becoming much more physically active.

If you are planning to become much more physically active than you are now, start by answering the seven questions in the box below. If you are between the ages of 15 and 69, the PAR-Q will tell you if you should check with your doctor before you start. If you are over 69 years of age, and you are not used to being very active, check with your doctor.

Common sense is your best guide when you answer these questions. Please read the questions carefully and answer each one honestly: check YES or NO.

### YES to one or more questions

Talk with your doctor by phone or in person BEFORE you start becoming much more physically active or BEFORE you have a fitness appraisal. Tell your doctor about the PAR-Q and which questions you answered YES.

- You may be able to do any activity you want — as long as you start slowly and build up gradually. Or, you may need to restrict your activities to those which are safe for you. Talk with your doctor about the kinds of activities you wish to participate in and follow his/her advice.
- Find out which community programs are safe and helpful for you.

### NO to all questions

If you answered NO honestly to all PAR-Q questions, you can be reasonably sure that you can:

- start becoming much more physically active — begin slowly and build up gradually. This is the safest and easiest way to go.
- take part in a fitness appraisal — this is an excellent way to determine your basic fitness so that you can plan the best way for you to live actively. It is also highly recommended that you have your blood pressure evaluated. If your reading is over 144/94, talk with your doctor before you start becoming much more physically active.

### Delay becoming much more active:

- if you are not feeling well because of a temporary illness such as a cold or a fever — wait until you feel better; or
- if you are or may be pregnant — talk to your doctor before you start becoming more active.

### PLEASE NOTE:

If your health changes so that you then answer YES to any of the above questions, tell your fitness or health professional. Ask whether you should change your physical activity plan.

Informed Use of the PAR-Q: The Canadian Society for Exercise Physiology, Health Canada, and their agents assume no liability for persons who undertake physical activity, and if in doubt after completing this questionnaire, consult your doctor prior to physical activity.

No changes permitted. You are encouraged to photocopy the PAR-Q but only if you use the entire form.

NOTE: If the PAR-Q is being given to a person before he or she participates in a physical activity program or a fitness appraisal, this section may be used for legal or administrative purposes.

"I have read, understood and completed this questionnaire. Any questions I had were answered to my full satisfaction."

NAME ____________________________________________

SIGNATURE ____________________________________________

DATE ____________________________________________

SIGNATURE OF PARENT or GUARDIAN (for participants under the age of majority)

WITNESS ____________________________________________

Note: This physical activity clearance is valid for a maximum of 12 months from the date it is completed and becomes invalid if your condition changes so that you would answer YES to any of the seven questions.
APPENDIX D

MEDICAL HISTORY QUESTIONNAIRE
MEDICAL HISTORY QUESTIONNAIRE

Full Name (printed): ___________________________  E-mail Address: ___________________________
Phone Number: (___) - __________  Date of birth: ___________________________

Directions: Please use Y (yes) or N (no) to indicate the appropriate response for each item.

PART I: KNOWN DISEASES

Do you currently have:
   ___ Cardiovascular disease, peripheral vascular disease, and/or cerebrovascular disease?
   ___ Asthma?
   ___ Interstitial lung disease?
   ___ Cystic fibrosis?
   ___ Chronic Obstructive Pulmonary Disease (COPD)?
   ___ Diabetes (Type 1 or 2)?
   ___ Any thyroid disorders?
   ___ Renal or liver disease?

PART II: SIGNS AND SYMPTOMS

___ Do you experience pain and/or discomfort in the chest, neck, jaw, arms, or other areas during mild exercise?
___ Do you feel short of breath at rest, with typical, daily activities, or with mild exercise?
___ Do you feel short of breath while lying down flat?
___ Are you awoken in the middle of night due to shortness of breath or severe coughing/wheezing?
___ Do you often feel dizzy at rest or with mild exercise?
___ Do you suddenly pass out or lose consciousness while at rest or with mild exercise?
___ Have you experienced ankle edema (swollen ankles)?
___ Do you have heart palpitations and/or tachycardia at rest or with mild exercise?
___ Do you suffer from muscle cramping, burning, numbness, or fatigue in your calf muscles at rest or with mild exercise?
___ Do you have a known heart murmur?
___ Do you have unusual fatigue with typical, daily activities?

(continued on next page)
PART III: CORONARY ARTERY DISEASE RISK FACTORS

___ Are you a male older than 45 years or a female older than 55 years?
___ Do you have a close blood relative who has had a heart attack or heart surgery before the age of 55 (Dad, Brother) or age 65 (Mom, Sister)?
___ Do you smoke, or did you just quit smoking within the past 6 months?
___ For the last 3 months, did you get less than 30 minutes of moderate-intense exercise, less than 3 days per week?
___ Are you at least 20lbs overweight?
___ Is your blood pressure over 140/90 mmHg, or are you on blood pressure medication?
___ Is your cholesterol greater than or equal to 200 mg/dL, or are you on cholesterol medication?
___ Is your fasting glucose greater than or equal to 100 ml/dL?

PART IV: MUSCULOSKELETAL CONDITIONS AND OTHER

___ Do you have musculoskeletal problems that limit what/how you exercise?
___ Have you had a major musculoskeletal injury (broken bones, torn ligaments/tendons, etc.) that has limited your ability to exercise in the past 12 months?
___ Do you have an implanted electrical device (e.g. pacemaker)?
___ Are you under the age of 18 or over the age of 40?
___ Do you have a tattoo that covers any of your posterior (backside of) wrist or forearm?

PART V: FEMALE HEALTH (females only)

___ Are you currently pregnant?
___ Are you actively trying to become pregnant?

PART VI: GENERAL SUPPLEMENTAL INFORMATION

___ Are you taking any prescription medications? (If so, please specify below)

<table>
<thead>
<tr>
<th>Name of drug/medication:</th>
<th>Reason for taking:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

___ Do you have any known allergies (food, latex, silicone, plants, animals, etc.)? (If so, please specify below)

Known Allergens:

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

RECRUITMENT SCRIPT
Dear [First Name],

Thank you for expressing interest in our research project which will assess wearable technology and the accuracy of physical activity trackers and smartwatches! This letter is to inform you that you have been invited to take part in our study. Dr. Amanda Salacinski, an Associate Professor for the Department of Kinesiology and Physical Education (KNPE) at Northern Illinois University (NIU) and William Kalamaras, a KNPE graduate student, will be the principle investigators in this research study. The aim of the study is to measure the accuracy of some of the health tracking capabilities found in the new Apple Watch and Microsoft Band. We would like to determine how accurate the functions of these devices are when compared to conventionally accepted tracking tools during exercise. You have been selected because you are a healthy, physically active individual, and meet the preliminary criteria to be involved as a participant.

Before you can be fully-cleared for participation in our study, you will be required to read and complete some paperwork for your safety. This paperwork will also ensure that you understand all of the requested duties and expectations of your participation, as well as your rights as a human participant in scientific research. Specifically, the attached Physical Activity Readiness Questionnaire (PAR-Q) and Medical History Questionnaire will be used to determine whether or not you are healthy enough to participate. Health-screening inclusion is determined through your ability to demonstrate ‘physical activity readiness’ by answering “no” to all questions included on the PAR-Q and receive a low cardiovascular disease (CVD) risk classification as determined by the American College of Sports Medicine’s (ACSM) guidelines. Additionally, female participants will need to clearly indicate that they are not currently pregnant, and do not plan to become pregnant on the Medical History Questionnaire in order to participate. Review and completion of the accompanied informed consent form will indicate that you understand the conditions and expectations of your participation, including any potential risks as a participant in research. Upon completion of this paperwork, you will then be asked to complete a preliminary assessment of your maximal exercise capacity by running on a treadmill. A percentage between 50% and 75% of your maximal value will then be used in three subsequent testing trials.

Your preliminary maximal exercise assessment will consist of running on a treadmill at increasing intensity until you can no longer continue. You should expect this preliminary assessment to take approximately 45 minutes in duration. Following this, you will be asked to schedule three testing sessions at least 36 hours apart from one another.

All three testing sessions will consist of you running on a treadmill without interruption for a total of 20 minutes. The sessions will start at a moderate exercise intensity (50% of your maximum), and increase to a higher intensity for the remaining sessions (75% of your maximum).

(continued on next page)
The intensity of any individual session is held constant for the entire trial. Before each testing session, you will have your weight, height, and body-fat measurements recorded. Throughout the exercise component of each session, you will be equipped with mask (to analyze your breath), a heart rate transmitter, an Apple Watch, and a Microsoft Band. Your body composition measurements will be taken using an InBody520. This device measures body fat, lean body mass, total body water, weight, and body mass index. To properly use this device, you will be asked to maintain a consistent diet on days of testing and to avoid the consumption of foods and large amounts of liquids for at least 1.5 hours prior to testing time. You will also be asked to refrain from consuming caffeine for at least 5 hours, and alcohol for at least 12 hours prior to testing time. Upon arrival, you will be asked to remain standing for 15 minutes before testing to ensure proper fluid balance. Upon completion in the study, this information will be available for you to obtain upon request.

Your preliminary assessment will take approximately 45 minutes, and the exercise portion of all three (3) sessions will take 20 minutes. We kindly ask that you allow approximately 35-40 minutes for each 20-minute session to accommodate for necessary preparatory protocols. In total, your participation in this study will take approximately 3 hours. We also ask that you please abstain from ingesting any additional stimulant and depressant drugs and/or supplements during testing. You will be asked to disclose any drugs, including prescription medications, prior to your participation in this research study.

All information will be kept confidential; your name will not be used on the final report or associated with data in any way. Only Northern Illinois University qualified research personnel will have access to the data and be present at the time of the study. All names and any other identifying information will be removed upon data entry. The risks from participating in this study include soreness, fatigue, physical exertion, and potential injury. Your participation in this study is completely voluntary. If for any reason you wish to discontinue, you have the right to withdraw at any time without penalty.

If you would like to schedule an initial assessment to participate, or if you have any further questions, please contact the graduate researcher, William Kalamaras at wkalamaras@niu.edu, or the thesis director and chair, Dr. Amanda Salacinski, at asalacinski@niu.edu. Questions about your rights as a research subject and any research related injury should be directed to the Research and Compliance Integrity Office at researchcompliance@niu.edu.

Thank you for your interest in our study.

Regards,
The Wearable Research Team
Department of Kinesiology and Physical Education
Northern Illinois University
APPENDIX F

RECRUITMENT FLYER
Wearable Technology Research Study
Apple Watch and Microsoft Band Validity Testing

RESEARCH PARTICIPANTS NEEDED TO TEST THE APPLE WATCH AND MICROSOFT BAND

The Department of Kinesiology and Physical Education (KNPE) is recruiting healthy/active male and female participants to participate in a research study to test the accuracy of health-tracking accuracy of two of the industry’s leading wearable devices.

To be eligible you must be:
- Regularly active
- Between the ages of 18 and 40 years
- Healthy / Free of Cardiovascular Disease
- Able to participate in physical activity

Eligible Participant’s will have the opportunity to use the Apple Watch and Microsoft Band during the duration of study.

Participants will be asked to perform one 45-minute pretest and three 20-minute testing sessions

Please allow an additional 15-20 minutes per session for preparation (~3 hours total)

For more information, please contact
Will Kalamaras wkalamaras@niu.edu
APPENDIX G

SUBJECT INFORMATION SHEET
## SUBJECT INFORMATION SHEET

** UNIQUE PARTICIPANT ID (UID):**

** DATE OF BIRTH:**

** GENDER:**
- ☐ MALE
- ☐ FEMALE

** AGE (YRS):**

## TESTING SESSION INFORMATION

<table>
<thead>
<tr>
<th>DATE</th>
<th>CONDITION</th>
<th>INTENSITY</th>
<th>VO₂</th>
<th>SPEED</th>
<th>ANTHROPOMETRIC DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRE</td>
<td>100%</td>
<td></td>
<td></td>
<td><strong>HEIGHT (CM, [IN]):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>ECW:TBW (RATIO):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BMI (KG·(M²)-1):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BODY FAT (%):</strong></td>
</tr>
<tr>
<td></td>
<td>T₁</td>
<td>50%</td>
<td></td>
<td></td>
<td><strong>HEIGHT (CM, [IN]):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>ECW:TBW (RATIO):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BMI (KG·(M²)-1):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BODY FAT (%):</strong></td>
</tr>
<tr>
<td></td>
<td>T₂</td>
<td>75%</td>
<td></td>
<td></td>
<td><strong>HEIGHT (CM, [IN]):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>ECW:TBW (RATIO):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BMI (KG·(M²)-1):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BODY FAT (%):</strong></td>
</tr>
<tr>
<td></td>
<td>T₃</td>
<td>75%</td>
<td></td>
<td></td>
<td><strong>HEIGHT (CM, [IN]):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>ECW:TBW (RATIO):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BMI (KG·(M²)-1):</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BODY FAT (%):</strong></td>
</tr>
</tbody>
</table>

## EQUIPMENT SIZING INFORMATION

### APPLE WATCH SPORTBAND

<table>
<thead>
<tr>
<th>SIZE:</th>
<th>Sm/Med</th>
<th>Med/LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITION:</td>
<td>WATCH</td>
<td>END</td>
</tr>
<tr>
<td>WATCH OS VERSION:</td>
<td>2.1 (13A441)</td>
<td></td>
</tr>
<tr>
<td>IOS VERSION:</td>
<td>9.0.2 (13A442)</td>
<td></td>
</tr>
</tbody>
</table>

### MICROSOFT BAND

<table>
<thead>
<tr>
<th>BAND SIZE:</th>
<th>Sm</th>
<th>Med</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASP POSITION:</td>
<td>CLOSED</td>
<td>END</td>
<td></td>
</tr>
<tr>
<td>F/W VERSION:</td>
<td>10.03.3304.0 09 R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEALTH VERSION:</td>
<td>1.3.20215.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### POLAR H7 SOFT STRAP

<table>
<thead>
<tr>
<th>STRAP SIZE:</th>
<th>XS-S</th>
<th>M-XXL</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITION:</td>
<td>CLOSED</td>
<td>END</td>
</tr>
</tbody>
</table>

### V2 MASK ASSAY

<table>
<thead>
<tr>
<th>FACE MASK SIZE:</th>
<th>PETITE</th>
<th>EXTRA-SMALL</th>
<th>SMALL</th>
<th>MEDIUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRAP SIZE:</td>
<td>SM</td>
<td>MED</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## SCHEDULE / NOTES

<table>
<thead>
<tr>
<th>SESSION</th>
<th>DATE</th>
<th>TIME</th>
<th>RESCHEDULES</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE</td>
<td>VO₂ MAX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T₁</td>
<td>50% VO₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T₂</td>
<td>75% VO₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T₃</td>
<td>75% VO₂</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX H

V₂ MASK SIZING GAUGE
HANS RUDOLPH inc.
MASK SIZING GAUGE

CHECK ADOBE PRINT SETTINGS BEFORE PRINTING
PAGE SCALING: NONE
IF PREVIEW IS OUTSIDE PRINTING BORDER
SELECT AUTO-ROTATE & CENTER

691143-C 10/2011

MASK SIZING GAUGE

Directions: Length measurements are used to determine the size of the Hans Rudolph Oro-Nasal mask that will best fit the patient. For measurement, the patient’s facial muscles should be relaxed and jaw closed. Place the Gauge tip to the deepest point of the nasal root depression (inion) and note the size indication where the bottom of the chin exits the gauge. The gray band widths indicate the size ranges.

This gauge is a reference to aid in sizing and not an exact measurement as other factors may need to be taken into account.

IF BETWEEN SIZES, WE RECOMMEND YOU USE THE SMALLER SIZE.

HANS RUDOLPH, inc.

PETITE (P)
EXTRA-SMALL (XS)
SMALL (S)
MEDIUM (M)
LARGE (L)

www.rudolphkc.com
APPENDIX I

MICROSOFT BAND SIZING GUIDE
The Microsoft Band comes in three sizes.
Follow these steps to choose the size that fits you best.

1. Print out the sizing guide.
   For the most accurate estimation, print the PDF at 100%.

2. Place the sizing guide on a flat surface. Align your arm with the arm on the sizing guide, and compare the height of your wrist with the purple bars. Note: don’t roll or wrap the paper.

3. Determine your best fit—small, medium, or large. If your wrist falls between two sizes, we recommend you choose the larger size for the most comfortable fit.
APPENDIX J

PARTICIPANT CHARACTERISTIC TABLES
<table>
<thead>
<tr>
<th>Age years</th>
<th>Height cm</th>
<th>VO$_{2\max}$ mL.kg$^{-1}$.min$^{-1}$</th>
<th>ACSM Ranking % rank</th>
<th>Body Mass kg</th>
<th>BMI$_{1}$ kg.$\mbox{~m}^{-2}$</th>
<th>Body Fat (%)</th>
<th>ECW: TBW$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>160</td>
<td>44.1</td>
<td>80.1</td>
<td>55.6</td>
<td>56.3</td>
<td>56.0</td>
<td>21.7</td>
</tr>
<tr>
<td>21</td>
<td>165</td>
<td>42.0</td>
<td>71.2</td>
<td>50.6</td>
<td>50.8</td>
<td>50.5</td>
<td>18.6</td>
</tr>
<tr>
<td>27</td>
<td>173</td>
<td>29.3</td>
<td>3.2</td>
<td>96.0</td>
<td>96.0</td>
<td>94.7</td>
<td>32.1</td>
</tr>
<tr>
<td>21</td>
<td>185</td>
<td>52.6</td>
<td>87.0</td>
<td>97.0</td>
<td>96.4</td>
<td>96.0</td>
<td>28.3</td>
</tr>
<tr>
<td>20</td>
<td>184</td>
<td>49.2</td>
<td>74.8</td>
<td>82.3</td>
<td>82.4</td>
<td>83.2</td>
<td>24.3</td>
</tr>
<tr>
<td>37</td>
<td>179</td>
<td>44.9</td>
<td>66.4</td>
<td>87.1</td>
<td>86.8</td>
<td>87.0</td>
<td>27.2</td>
</tr>
<tr>
<td>25</td>
<td>195</td>
<td>61.6</td>
<td>98.4</td>
<td>98.8</td>
<td>97.5</td>
<td>96.6</td>
<td>26.0</td>
</tr>
<tr>
<td>20</td>
<td>184</td>
<td>50.5</td>
<td>80.0</td>
<td>68.0</td>
<td>68.4</td>
<td>68.9</td>
<td>19.9</td>
</tr>
<tr>
<td>23</td>
<td>168</td>
<td>50.4</td>
<td>81.1</td>
<td>75.3</td>
<td>75.8</td>
<td>76.0</td>
<td>26.7</td>
</tr>
<tr>
<td>20</td>
<td>167</td>
<td>51.3</td>
<td>82.9</td>
<td>74.1</td>
<td>73.4</td>
<td>73.7</td>
<td>26.7</td>
</tr>
<tr>
<td>22</td>
<td>185</td>
<td>52.2</td>
<td>86.2</td>
<td>76.5</td>
<td>75.9</td>
<td>76.3</td>
<td>22.4</td>
</tr>
<tr>
<td>19</td>
<td>170</td>
<td>43.2</td>
<td>42.4</td>
<td>90.1</td>
<td>88.3</td>
<td>88.6</td>
<td>31.2</td>
</tr>
<tr>
<td>24</td>
<td>185</td>
<td>40.9</td>
<td>32.4</td>
<td>80.9</td>
<td>80.4</td>
<td>79.8</td>
<td>23.6</td>
</tr>
<tr>
<td>20</td>
<td>169</td>
<td>45.5</td>
<td>56.1</td>
<td>66.8</td>
<td>67.7</td>
<td>65.3</td>
<td>23.4</td>
</tr>
<tr>
<td>22</td>
<td>164</td>
<td>34.4</td>
<td>28.6</td>
<td>53.0</td>
<td>54.0</td>
<td>54.0</td>
<td>19.8</td>
</tr>
<tr>
<td>21</td>
<td>175</td>
<td>41.7</td>
<td>69.6</td>
<td>54.2</td>
<td>54.7</td>
<td>54.3</td>
<td>17.7</td>
</tr>
<tr>
<td>20</td>
<td>163</td>
<td>38.4</td>
<td>50.7</td>
<td>72.2</td>
<td>71.7</td>
<td>70.8</td>
<td>27.2</td>
</tr>
<tr>
<td>26</td>
<td>175</td>
<td>50.7</td>
<td>83.4</td>
<td>87.6</td>
<td>87.5</td>
<td>86.6</td>
<td>28.6</td>
</tr>
<tr>
<td>23</td>
<td>161</td>
<td>49.8</td>
<td>94.5</td>
<td>68.1</td>
<td>68.3</td>
<td>69.5</td>
<td>26.3</td>
</tr>
<tr>
<td>25</td>
<td>177</td>
<td>50.2</td>
<td>81.3</td>
<td>80.8</td>
<td>81.7</td>
<td>81.4</td>
<td>25.8</td>
</tr>
<tr>
<td>23</td>
<td>168</td>
<td>37.6</td>
<td>46.6</td>
<td>62.1</td>
<td>62.4</td>
<td>61.4</td>
<td>22.0</td>
</tr>
<tr>
<td>21</td>
<td>179</td>
<td>56.5</td>
<td>94.3</td>
<td>88.8</td>
<td>87.8</td>
<td>88.2</td>
<td>27.7</td>
</tr>
<tr>
<td>22</td>
<td>182</td>
<td>55.5</td>
<td>93.1</td>
<td>81.5</td>
<td>82.1</td>
<td>81.9</td>
<td>24.6</td>
</tr>
</tbody>
</table>
Table 8
Participant Descriptive Statistics by Session (Overview)

|                  | Age (years) | Height (cm) | VO\textsubscript{2,max} (mL kg\textsuperscript{-1} min\textsuperscript{-1}) | ACSM Ranking | Body Mass (kg) | BMI\textsubscript{1} (kg m\textsuperscript{-2}) | Body Fat (%) | ECW: TBW\textsubscript{2} | T1 | T2 | T3 | T1 | T2 | T3 | T1 | T2 | T3 | T1 | T2 | T3 |
|------------------|-------------|-------------|----------------------------------------------------------------|--------------|---------------|---------------------------------------------|--------------|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Mean             | 22.7        | 174.5       | 46.6                                                          | 68.9         | 76.0          | 75.9                                        | 75.7         | 24.9                        | 24.9            | 24.8            | 17.2            | 17.3            | 17.4            | 0.368           | 0.368           | 0.368           |
| Median           | 22.0        | 175         | 49.2                                                          | 80.0         | 76.5          | 75.9                                        | 76.3         | 25.8                        | 25.6            | 25.4            | 14.2            | 14.4            | 15.3            | 0.368           | 0.366           | 0.366           |
| Std Dev          | 3.8         | 9.4         | 7.6                                                           | 24.7         | 14.4          | 14.0                                        | 14.0         | 3.8                         | 3.7             | 3.7             | 8.0             | 8.2             | 8.0             | 0.007           | 0.007           | 0.007           |
| Min              | 19          | 160         | 29.3                                                          | 3.2          | 50.6          | 50.8                                        | 50.5         | 17.7                        | 17.9            | 17.7            | 6.9             | 6.5             | 6.3             | 0.356           | 0.355           | 0.355           |
| Max              | 37          | 195         | 61.6                                                          | 98.4         | 98.8          | 97.5                                        | 96.6         | 32.1                        | 32.1            | 31.6            | 35.8            | 36.3            | 34.8            | 0.381           | 0.386           | 0.384           |

N=23 (16 males, 7 females)
1Body Mass Index
2Extracellular Water: Total Body Water
APPENDIX K

CONCORDANCE CORRELATION COEFFICIENTS
Table 9
Concordance Correlation Coefficients ($r_c$) with 95% Confidence Intervals (CI) for Heart Rate and Energy Expenditure at 50% and 75% VO$_2$max

<table>
<thead>
<tr>
<th>Intensity (Trial)</th>
<th>Heart Rate (HR)</th>
<th>Energy Expenditure (EE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Polar H7 HRM vs. 42mm Apple Watch</td>
<td>Polar H7 HRM vs. Microsoft Band</td>
</tr>
<tr>
<td>50% VO$_2$max ($T_1$)</td>
<td>$r_c$ 1.00 (0.99 – 1.00)</td>
<td>0.81 (0.64 – 0.91)</td>
</tr>
<tr>
<td>75% VO$_2$max ($T_2$)</td>
<td>$r_c$ 0.99 (0.98 – 1.00)</td>
<td>0.65 (0.43 – 0.79)</td>
</tr>
</tbody>
</table>

N=23 (16 males, 7 females)