PSYCHOMETRIC PARAMETERS OF ZEPHYR BIOHARNESS AND FITBIT CHARGE

RELIABILITY, VALIDITY AND AGREEMENT PARAMETERS OF ZEPHYR BIOHANRESS AND FITBIT CHARGE, MEASURES OF HEART RATE AND ACTIVITY AT REST, DURING THE MODIFIED CANADIAN AEROBIC FITNESS TEST AND RECOVERY.

By

Goris Nazari

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

In Rehabilitation Sciences

McMaster University

MASTER OF SCIENCE (2016) McMaster University,

(Rehabilitation Science) Ontario, Canada

TITLE: Reliability, validity and agreement parameters of Zephyr Bioharness and Fitbit Charge, measures of heart rate and activity at rest, during the Modified Canadian Aerobic Fitness Test and recovery.

AUTHOR: Goris Nazari BSc. P.T.

SUPERVISOR: Dr. Joy C. MacDermid.

Supervisory Committee: Dr. Julie Richardson

Dr. Ada Tang

Dr. Kathryn Sinden

NUMBER OF PAGE x, 162

CONTRIBUTIONS

Goris Nazari was responsible for developing the research questions, the study design, ethics submissions and collection, analysis and interpretation of data as well as the drafting of the manuscripts and incorporating feedback. Dr. Joy C. MacDermid was responsible for reviewing and refining of the research questions and study design. Dr. Joy MacDermid assisted in data interpretation and also provided editorial assistance with manuscript preparation. Dr. Kathryn E. Sinden, Dr. Julie Richardson and Dr. Ada Tang provided feedback during the committee meetings on interpretation of data and reviewed the manuscripts.

ABSTRACT

Introduction: Technological innovations have lead to the development of Wearable Physiological Monitoring devices, that have enabled researchers and clinicians in real-time monitoring of physiologic function within a field setting. However, it is important to establish the psychometric properties of a device prior to its utilisation.

Thesis Objectives: A systematic review was conducted to provide a summary and appraise the quality of the literature on psychometric parameters of Zephyr Bioharness and Fitbit devices. Based on this review, we addressed the current gaps in the literature regarding the reliability parameters of Zephyr Bioharness and Fitbit Charge devices, and established the validity and agreement properties of Fitbit Charge device.

Methods: For our systematic review, we searched the Google Scholar and PubMed databases to identify articles. To establish the reliability, validity and agreement parameters of Zephyr Bioharness and Fitbit Charge devices, a convenience and snowball sampling approaches were used to recruit sixty participants (30 females) from university student, staff, faculty population, and MacSeniors Community Program at McMaster University. The performance of Zephyr and Fitbit devices were assessed throughout three phases; rest, Modified Canadian Aerobic Fitness Test and recovery.

Results: In our study, at rest, inter-session average heart rate (beats/min.) ICCs (SEM) for Zephyr and Fitbit ranged from 0.90 – 0.94 (1.73 – 2.37) and 0.88 – 0.94 (1.83 – 2.67) respectively. At mCAFT, the Zephyr ICCs (SEM) ranged from 0.91 – 0.97 (3.12 – 4.64) and 0.85 – 0.98 (3.28 – 4.88) for the Fitbit. Throughout the recovery, the ICCs (SEM) ranged from 0.93 – 0.97 (2.65 – 4.66) and 0.76 – 0.91 (3.17 – 4.67) for Zephyr and Fitbit

devices respectively. Pearson's correlation coefficients and (Mean differences) for heart rate variable were 0.97 – 0.99 (-0.60 – 0.02) at Rest, 0.89 – 0.99 (13.51 – 0.62) at submaximal testing and 0.70 – 0.84 (-0.54 – 2.52) throughout recovery. The average agreement bias of heart rate in pair-wise device comparison indicated mean differences of -0.20, 4.00 and 1.00 at rest, sub-maximal testing and recovery respectively.

Conclusions: We identified fair to very good quality evidence from 14 studies. The Zephyr Bioharness and Fitbit Charge devices demonstrated excellent reliability measures, and the Fitbit Charge device heart rate variable demonstrated strong to very strong correlations when concurrently compared with Zephyr, and provided valuable information regarding its interchangeable use in a sample of sixty healthy male and female participants of various age groups during a resting, standardized submaximal fitness and recovery phases.

ACKNOWLEDGEMENTS

My thesis supervisor, Dr. Joy MacDermid has been a phenomenal source of support and guidance. She is an absolutely amazing mentor and leader, one that I aspire to become. Words simply can not express how grateful I am to you.

I would also like to express my gratitude to Dr. Kathryn Sinden for giving me thoughtful and comprehensive feedback on my writings so that I could express my ideas much more clearly. I am also thankful to the rest of my committee members Dr. Julie Richardson and Dr. Ada Tang for their support at various levels of my thesis project.

I would also like to thank Margaret Lomotan at McMaster University for her continuous operational support of my research.

A special thank-you to my parents. My father, Serj Nazari, who taught me to be so determined and fearless no matter what obstacle is in my path, and my mother, Gilda S. Nazari, for her unconditional love and immense belief.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Wearable Devices:

Technological innovations have lead to the development of Wearable Physiological Monitoring (WPM) devices, that have enabled researchers and clinicians in real-time monitoring of physiologic function within a field setting (Li et al., 2016). WPM devices are small, portable, less costly and use wireless technology to provide a non-invasive long-term method of capturing physiological measures on the wearer in the home and community settings (Johnstone et al., 2012; Bonato 2010). Prior to the development of WPM devices, manual and observational methods were used to quantify physiological measures such as heart rate, respiratory rate and number of steps taken in a field setting (Bianchi et al., 2013). In addition, levels of energy expenditures were assessed using subjective measures including questionnaires or surveys (Ceesay et al., 1989). Manual pulse palpation of an artery and direct observation of respiration are costly in terms of time and personnel and more importantly provide inaccurate results in assessing heart rate and respiratory rate measures respectively (Bianchi et al., 2013; Jovanov et al., 2002). Furthermore, in assessing energy expenditures through subjective measures, issues such as under or over estimation of activity levels due to social desirability or intricate questionnaires, as well as inaccurate reporting due to impaired memory among young or elderly populations have been reported (Vanheesa et al., 2005).

Importance of Physiologic Measures:

Monitoring of physiological measures such as; heart rate, steps taken and energy expenditures, have important implications. Long-term studies have reported that higher

levels of resting heart rate, lower maximal exercise-induced heart rate and a delay in heart rate recovery after exercise, have been found to be good predictors of cardiovascular and all-cause mortality among health individuals regardless of physical fitness levels, age or conventional coronary risk factors (Sandvik et al., 1995; Cole et al., 1999; Jensen et al., 2010; Jensen et al., 2013). Furthermore, a twenty-three year follow up study of 5713 individuals with no history of cardiovascular disease, demonstrated that increased risk of sudden death was associated with a resting heart rate of greater that 75 beats per minute, an increase in heart rate during exercise that was less than 89 beats per minute (peak exercise levels – resting levels) and a decrease in heart rate of less that 25 beats per minute after cessation of exercise (peak exercise level – one minute post exercise) (Jouven et al. 2005).

To determine the effectiveness of an exercise session, it is necessary to evaluate its intensity levels by calculating the total sum of energy expended throughout the entire exercising session. This involves the use of indirect calorimetry, – a laboratory-based method, to measure the amount of gaseous exchange and energy produced. For every litre of oxygen consumed and carbon dioxide produced (gaseous exchange), a specific amount of energy is expended (kilocalories) depending on individuals' body weight as well as the exercise/task undertaken (Achten & Jeukendrup 2003; Hills, Mokhtar & Byrne 2014). The application of this method is limited and can not be extended to non-laboratory settings. A much more practical and valid method of assessing exercise intensity levels in a field setting involves the use of heart rate monitoring (Achten & Jeukendrup 2003). A linear relationship exists between heart rate and oxygen consumption (VO_2) levels, and since VO_2

is regarded as a good indicator of energy expenditures, heart rate monitoring can be used to precisely estimate exercise intensity levels (Achten & Jeukendrup 2003).

Maximal oxygen consumption ($VO_{2\text{-}max}$), reflects upon an individual's cardiorespiratory fitness (CRF) levels (American College of Sports Medicine 2013). CRF is a term that relates to an individual's ability to carry out dynamic, moderate -to-high intensity exercise involving large muscle groups for extended periods. The importance of measuring CRF is that it is regarded as one of the best indicators of collective health, which provides both prognostics and diagnostic information regardless of presence or absence of chronic disease (American College of Sports Medicine 2013). The evidence reports that a low level of cardiorespiratory fitness is considered as an independent risk factor for all-cause and cardiovascular mortality. In addition, CRF is associated with an one's functional capacity. Therefore, an individual with higher CRF levels would be able to carry out occupational tasks and activities of daily living with little effort and minimal fatigue as compared to one with lower CRF levels (American College of Sports Medicine 2013). Throughout a large range of exercise intensities and on group levels, monitoring of heart rate can be used to estimate $VO_{2\text{-}max}$, since a linear relationship exists between heart rate and VO_2 levels (Achten & Jeukendrup 2003).

Several studies have concluded that increased work related physiological demands can have a negative impact on worker's ability to perform muscular task, lower their attentiveness and ultimately lead to an increased risk of injuries (Garet et al., 2005; Bouchard & Trudeau 2008; Rwamamara et al., 2010; Hsie et al. 2010). Heart rate monitoring is regarded as a very promising physiological parameter in assessing workers'

physiological demands in a field setting (Kirk et al., 2001; Bussmann et al., 2004; Garet et al., 2005).

The contraction of human skeletal muscles to produce bodily movements, that ultimately lead to energy expenditure, ranging from low to high, are elements that define physical activity (Caspersen, Powell & Christenson 1985). Physical activity is regarded as an effective preventive tool for various health conditions irrespective of age, gender, ethnicity or socioeconomic status (Tremblay et al., 2011). The current literature reports a strong inverse association between physical activity and prevalence of various morbidities. Health conditions such as obesity, type 2 diabetes mellitus, hypertension and metabolic syndrome are more prevalent among those with insufficient levels of physical activity compared to those with higher levels (Paffenbarger et al., 1983; Helmrich et al. 1991; Jebb et al. 1999; Lahti-Koski et al., 2002; Kriska et al. 2003; Rennie et al., 2003; Hu et al., 2004; Dunstan et al., 2005). In addition, a considerable increase in cardiovascular mortality has been observed with lower levels of physical activity (Smith et al., 2000). When the greater part of the population has a level of physical activity less than the recommended levels, there is a burden on the Canadian health care system due to associated health costs. In 2009, insufficient physical activity accounted for $ 6.70 billion as direct health care costs; hospital, drug and physician care expenditures, and indirect expenditures; economic output lost due to illness, injury-related work disability or premature death (Janssen 2012). In an attempt to promote regular physical activity, small, easy to use and relatively inexpensive ‘Accelerometers’ have been introduced that are usually mounted on the waist or worn on the wrist to provide continuous recording of data for several days/weeks. Accelerometers

are used to register accelerations and decelerations produced by body movements. Accelerometers use piezoelectric transducers and microprocessors to record the magnitude of acceleration in three (vertical, medio-lateral and anterior-posterior) dimensions, and then quantify acceleration (which is proportional to external forces), into a digital dimensionless "counts" (Bouten et al., 1994; Vanheesa et al., 2005). With its ability to monitor movement in vertical, medio-lateral and anterior-posterior dimensions, accelerometers are used to count the total number of steps taken in one day – upon waking up till going to sleep (Vanheesa et al., 2005). The total number of steps taken in one day, are then used to categorise individuals into sedentary (< 5000 steps), low active (5000 – 7499 steps), somewhat active (7500 – 9999 steps), active (10,000 – 12,500) and highly active (> 12,500 steps) physical activity levels (Hills, Mokhtar & Byrne 2014).

As stated previously, accelerometers record the magnitude of body accelerations and decelerations into "counts". These accelerometry counts have demonstrated linear relationships with energy expenditures (Bouten et al., 1994). Therefore, energy expenditures during physical activity can be estimated using linear regression equations when variables; height, body weight, age and gender, are calculated (Bouten et al., 1994).

Review of Wearable Physiological Monitoring devices:

Since the past several years, numerous Wearable Physiological Monitoring (WPM) devices have been introduced to the market (Yang & Hsu 2010; Collier 2015). To provide an outlook on their developments, the most commonly used chest, waist and wrist-based WPM devices in research literature have been displayed (Table – 1.1).

Zephyr BioHarness^{TM} (Zephyr Technology Corporation, Annapolis, MD, US), is a U.S. FDA-approved wireless WPM device capable of recording and transmission of physiological measures such as; heart rate, heart rate variability, electrocardiogram signals, respiratory rate, estimated core temperature, posture and activity levels (Zephyr Technology Manual 2012). The device measures heart rate and electrocardiogram signals through recording of cardiac electrical impulses by conductive fabric skin electrodes and reports beats per minute ($b \cdot min^{-1}$) and electrocardiogram amplitude in milli-volt (mV) over time, respectively. Respiratory rate is measured through a size differential determined by a strap sensor under the wearer's right arm during expansion and contraction of thoracic cavity producing an output as breaths per minute ($br \cdot min^{-1}$) (Zephyr Technology Manual 2012). This differential along with cardiac impulses are sent to and processed by the BioModule. The module also consists of an internal thermistor to estimate temperature as well as an inner 3-axis accelerometer which uses piezoelectric technology to determine the subject's activity levels in terms of acceleration. The acceleration data is established in gravitational force g (i.e. m/s^2) as a Vector Magnitude Unit (VMU) which is the square root of sum of the squares of vertical, sagittal and lateral axes. Posture also uses piezoelectric technology. It functions as an inclinometer and captures data in angular degrees (^{o}), ranging from -180 to +180, detecting how far the BioHarness deviates from the vertical plane (y-axis) "off the vertical". The BioHarness consists of an adjustable chest belt featuring conductive fabric sensors and an electronic BioModule that snaps on to the belt (Zephyr Technology 2012). The device is fitted comfortably on the chest at the lower sternum for both men and women and weighs 85grams. It can capture data for 26 hours upon a

successful three-hour charge and can log up to 20 days of data. The Zephyr BioHarness device has been developed to improve convenience, safety, quality of health and performance. It offers a field-based monitoring of physiologic measures and can simultaneously record heart rate, heart rate variability, electrocardiogram signals, respiratory rate, estimated core temperature, posture and activity levels unobtrusively (Johnstone et al., 2012). The device provides a unique advantage of processing physiological data in physically demanding environments (-30 – +60 °C), and when real-time long range (300 meters) high level wireless monitoring for extended periods of time are required (Zephyr Technology Manual 2012).

Fitbit Charge Heart Rate (Fitbit Inc., San Francisco, CA, US) is a wrist-based WPM device, made of a flexible, elastomer wristband with an operating temperature that ranges from -20 – +45 °C. It is the first wrist-based WPM device with an optical heart rate monitor and PurePulse^{TM} LED lights that reflect on the wearer's skin to detect blood volume changes, and then use a set of sophisticated algorithms to provide continuous automatic resting and exercise heart rate measures (Fitbit Charge HR 2015). In addition, it provides "on-device" feedback display, while monitoring and recording heart rate measures at 1-second intervals while exercising, and 5-second intervals during all other times. Fitbit Charge Heart Rate device also includes 3-axis accelerometry sensors capable of capturing the total number of steps taken, energy expenditures, distance travelled, floors climbed and sleep time (Fitbit Charge HR 2015). The device also synchronises with personal computers, is sweat, rain and splash proof and can store data up to 30 days that would last for 5-days upon a two-hour charge.

Psychometric Properties:

It is important to establish the psychometric properties of a device prior to its utilisation (Streiner & Norman 1995). The reliability of a device quantifies its random measurement error and are explained in relative and absolute terms (McDowell & Newell 1996). Relative reliability is concerned with the ability of a device to differentiate among the individual's being assessed (MacDermid et al., 2009). It is a variance ratio and is influenced by the amount of variability within a sample. Absolute reliability of a device is the degree to which similar results are achieved following repeated measures using the same device within a stable condition (MacDermid et al., 2009). Validity of a device implies the extent to which a device measures what it is intended to measure (Brazier & Deverill 1999). Concurrent validity is achieved when simultaneous recordings are made by a device (to be validated) and its criterion (Portney & Watkins 1993). However, to assess and detect change over time, the responsiveness of the device must also be established (Wright & Young 1998). In addition, the advantages of establishing levels of agreement among devices, enables researchers and clinicians to assess the degree to which a new device differs from a criterion (gold standard) measure, and whether the devices can be used interchangeably or the new device can replace the gold standard measure (Bland & Altman 1986).

Importance of A Systematic Review:

Studies concerning the psychometric properties of devices, assess a wide range of measures and may not address all the reliability, validity and responsiveness aspects of

psychometric properties (MacDermid et al., 2009). It is useful to summarise and appraise the quality of the existing psychometric literature on a given device within standardised domains, so that this knowledge can provide some clinical recommendations regarding its use. A systematic review involves extraction of specific information (corresponding to the psychometric properties) of a given device, and critical appraisal of the study design to evaluate the value of the knowledge presented in an article using a structured clinical measurement specific appraisal tool. (MacDermid et al., 2009).

Composition of Dissertation Papers:

The dissertation is compromised of three papers (Chapters 2 – 4). The papers include a systematic review paper and two research studies that were completed as part of candidate's MSc program requirements in School of Rehabilitation Sciences at McMaster University.

A systematic review paper (Chapter – 2) provides a summary and quality appraisal of the psychometric parameters reported on Zephyr Bioharness, Fitbit Charge and Fitbit-ONE devices. The Fitbit-ONE device was included in this paper, because it would be concurrently compared with Fitbit Charge Heart Rate device in Chapter – 4, therefore, articles on psychometrics of Fitbit-ONE were also appraised for quality. Based upon this systematic review, we synthesized, appraised the quality and identified the gaps in the current literature concerning the psychometric parameters of Zephyr Bioharness and Fitbit devices, and addressed the existing gaps in Chapters 3 – 4. The second research paper (Chapter – 3), establishes the reliability parameters of Zephyr Bioharness and Fitbit Charge

Heart Rate devices at rest, during a sub-maximal activity and throughout recovery, among a sample of healthy males and females. In this paper the heart rate, number of steps taken and energy expenditure variables were assessed. The third research paper (Chapter – 4) establishes the validity and inter-instrument agreement parameters between Zephyr Bioharness and Fitbit Charge Heart Rate devices. In this paper, the Fitbit Charge device is concurrently compared with Zephyr Bioharness for the heart rate variable, and Fitbit-ONE for number of steps taken and energy expenditure variables.

In summary, research in this dissertation attempts to address the psychometric literature gaps, by assessing the reliability, validity and agreement parameters of Zephyr Bioharness and Fitbit Charge Heart Rate devices.

References:

Achten, J., and Jeukendrup, A. E., (2003). Heart Rate Monitoring Applications and Limitations. *Sports Med*, 33 (7): 517-538.

American College of Sports Medicine (2013). Health-Related Physical Fitness Assessment Manual. Fourth edition, Philadelphia, PA.

Bianchi W., Dugas A. F., Hsieh Y. H., Saheed M., Hill P., Lindauer C., …Rothman R. E. (2013). Revitalizing a vital sign: Improving detection of tachypnea at primary triage. *Annals of Emergency Medicine*, 61 (1), 37 – 43.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327 (8476), 307–10.

Bonato, P., (2010). Advances in Wearable Technology and its Medical Applications. *Conf. Proc IEEE Eng. Med Biol. Soc.*, 2021-2024. Buenos Aires, Argentina.

Bouchard, D. R., and Trudeau, F., (2008). Estimation of energy expenditure in a work environment: comparison of accelerometry and oxygen consumption/heart rate regression *Ergonomics*, 51, 663–670.

Bouten, C.V., Westerterp, K.R., Verduin, M., and Janssen, J.D. (1994). Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Med Sci Sports Exerc.*, 26, 1516–1523.

Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related qualityof life: learning from psychometrics. *Health Economics*, 8, 41–51.

Bussmann, J.B., Van den Berg-Emons, H.J.G., Angulo, S.M., Stijnen, T., and Stam, H.J., (2004). Sensitivity and reproducibility of accelerometry and heart rate in physical strain assessment during prosthetic gait. *Eur. J. Appl. Physiol*., 91, 71–78

Caspersen C.J., Powell K.E., and Christenson G.M. (1985). Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research. *Public Health Reports*, 100(2) 126 – 131.

Ceesay, S. M., Prentice, A.M., Day, K.C., Murgatroyd, P.R., Goldberg, G.R., Scott, W., and Spurr, G.B., (1989). The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br J Nutr*, 61(2):175–86.

Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E., and Lauer M. S. (1999). Heart-rate recovery immediately after exercise as a predictor of mortality. *The New England Journal of Medicine,* 341(18), 1351 – 1357.

Collier, R. (2015). Wearable Technologies for Healthcare Innovation. Wearable Technologies for Healthcare Innovation.

Dunstan, D.W., Salmon, J., Owen, N., Armstrong, T., Zimmet, P.Z., Welborn, T.A., … AusDiab Steering Committee (2005). Associations of TV viewing and physical activity with the metabolic syndrome in Australian adults. *Diabetologia*, 48, 2254–2261.

Fitbit charge HR Wireless Heart Rate + Activity Wrist Band (2015, January*). Fitbit charge HR$_{TM}$ Product Manual Version 1.0.* Retrieved from https://www.fitbit.com/ca/chargehr

Garet, M., Boudet, G., Montaurier, C., Vermorel, M., Coudert, J., and Chamoux, A., (2005). Estimating relative physical workload using heart rate monitoring: a validation by whole-body indirect calorimetry. *Eur. J. Appl. Physiol.,* 94, 46–53

Helmrich, S.P., Ragland, D.R., Leung, R.W., and Paffenbarger, R.S. (1991). Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med*, 325, 147–152.

Hills, A. P., Mokhtar, N., and Byrne, N. M., (2014). Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in Nutrition*, 1(5).

Hsie, M., Hsiao, W., Cheng, T., and Chen, H., (2009). A model used in creating a work-rest schedule for laborers *Autom. Constr.*, 18, 762–769

Hu, G., Barengo, N.C., Tuomilehto, J., Lakka, T.A., Nissinen, A., and Jousilahti, P., (2004). Relationship of physical activity and body mass index to the risk of hypertension: A prospective study in Finland. *Hypertension*, 43, 25–30.

Janssen I. (2012). Health care costs of physical inactivity in Canadian adults. *Appl. Physiol. Nutr. Metab.*, 37: 803 – 806.

Jebb, S.A., and Moore, M.S. (1999). Contribution of a sedentary lifestyle and inactivity to the etiology of overweight and obesity: Current evidence and research issues. *Med Sci Sports Exerc*, 31, 534–541.

Jensen, M. T., Marott, J. L., Allin, K. H., Nordestgaard, B. G., and Jensen, G. B. (2010). *European Journal of Preventive Cardiology*, 19(1), 102 – 108.

Jensen, M. T., Suadicani, P., Hein, H. O., and Gyntelberg, F., (2013). Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the Copenhagen Male Study. *Heart,* 99, 882–887.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012a). Bioharness^{TM} multivariable monitoring device: Part I: Validity. *Journal of Sports Science and Medicine* 11(3), 400- 408.

Jouven, X., Empana, J. P., Schwartz, P. J., Desnos, M., Courbon, D., and Ducimetière, P., (2005). Heart-Rate Profile during Exercise as a Predictor of Sudden Death. *The New England Journal of Medicine,* 352(19), 1951-1958.

Jovanov E., O'Donnel A., Morgan A., Priddy B. & Hormigo R. (2002). Prolonged telemetric monitoring of the heart rate variability using wireless intelligent sensors and mobile getaway. Proceedings of the Second Joint *Eng. in Med. and Bio. Soc. / Biomed. Eng. Soc.* Conference, Houston, TX, USA.

Kirk, P.M., and Sullman, M.J., (2001). Heart rate strain in cable hauler choker setters in New Zealand logging operations. *Appl. Ergon.*, 32, 389–398

Kriska, A.M., Saremi, A., Hanson, R.L., Bennett, P.H., Kobes, S., Williams, D.E., and Knowler, W. C., (2003). Physical activity, obesity, and the incidence of type 2 diabetes in a high-risk population. *Am J Epidemiol*, 158: 669– 675.

Lahti-Koski, M., Pietinen, P., Heliövaara, M., and Vartiainen, E. (2002). Associations of body mass index and obesity with physical activity, food choices, alcohol intake, and smoking in the 1982–1997 FINRISK studies. *Am J Clin Nutr*, 75, 809–817.

Li, R.T., Kling, S. R., Salata, M. J., Cupp S. A., Sheehan J., and Voos J. E., (2016). Wearable Performance Devices in Sports Medicine. *Sports Health*, 8(1), 74 – 78.

MacDermid, J. C., Walton, D. M., Avery, S., Blanchard, A., Etruw, E., Mcalpine, C., and Goldsmith, C. H. (2009). Measurement Properties of the Neck Disability Index: A Systematic Review. *Journal of orthopaedic & sports physical therapy*, 39(5), 400 – 417.

McDowell, I., and Newell, C. (1996). Measuring Health. A Guide to Rating Scales and Questionnaires. 2nd ed. New York: Oxford University Press.

Paffenbarger, R.S., Wing, A.L., Hyde, R.T., and Jung, D.L. (1983). Physical activity and incidence of hypertension in college alumni. *Am J Epidemiol*, 117: 245–257.

Portney L.G., and Watkins M.P. (1993). Foundations of clinical research: applications to practice. Connecticut: Appleton & Lange.

Rennie, K., McCarthy, N., Yazdgerdi, S., Marmot, M., and Brunner, E. (2003). Association of the metabolic syndrome with both vigorous and moderate physical activity. *Int J Epidemiol*, 32: 600–606.

Rwamamara, R. A., Lagerqvist, O., Olofsson, T., Johansson, B. M., and Kaminskas, K. A., (2010). Evidence-based prevention of work-related musculoskeletal injuries in construction industry. *J. Civ. Eng. Manag.,* 16, 499–509

Sandvik, L., Erikssen, J., Ellestad, M., Erikssen, G., Thaulow, E., Mundal, R., and Rodahl, K., (1995). Heart rate increase and maximal heart rate during exercise as predictor of cardiovascular mortality: 16-year follow up study of 1960 health men. *Coronary Artery Disease*, 6, 667-678.

Smith, G.D., Shipley, M.J., Batty, G.D., Morris, J.N., and Marmot, M. (2000). Physical activity and cause-specific mortality in the Whitehall study. *Public Health*, 114: 308–315.

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: a practical guide to their development and use* (2nded). New York: Oxford University Press.

Tremblay, M. S., Warburton, D. E. R., Janssen, I., Paterson, D. H., Latimer, A. E., Rhodes, R. E., … Dugga, M. (2011). New Canadian Physical Activity Guidelines. *Appl. Physiol. Nutr. Metab*. 36: 36–46.

Vanheesa, L., Lefevre, J., Philippaerts, R., Martens, M., Huygens, W., Troosters, T., and Beunen, G., (2005). How to assess physical activity? How to assess physical fitness? *European Journal of Cardiovascular Prevention and Rehabilitation*, 12(2), 102-114.

Wright J.G, and Young N.L. (1998). A comparison of different indices of responsiveness. *J Clin Epidemiol*, 50(3):239–46.

Yang, C., and Hsu, Y., (2010). A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, 10, 7772-7788

Zephyr Technology (2012, September 12). *BioHarness 3.0 User ManualBioHarness 3.0 User Manual*. Retrieved from http://zephyranywhere.com/products/bioharness-3/

Table – 1.1 Wearable Physiological Monitoring device specification comparison.

	Zephyr Bio-Harness	**Polar H7**	**Wahoo TICKR X**	**Viiiiva 4iiii**	**Garmin HRM**	**Fitbit Charge-HR**
Release Date	July 2011	June 2012	Jan. 2015	June 2013	Sep. 2015	Jan. 2015
Device Placement	Chest	Chest	Chest	Chest	Chest	Wrist
Weight (g)	85	158	55	-	59	-
Operating Temperature	-30 – +60 °C	-10 – +50 °C	-	0 – +50 °C	-	-20 – +45 °C
Accelerometer Type	Piezoelectric	n/a	Piezoelectric	n/a	Piezoelectric	Piezoelectric
Accelerometer axis	3	n/a	3	n/a	3	3
Sampling Frequency	2.40 GHz	2.40 GHz	2.40 GHz	2.40 GHz	2.40 GHz	-
Battery Type	4.2 V Lithium	3V Lithium	3V Lithium	3V Lithium	3V Lithium	Lithium-polymer
Battery Life	26 hours	350 hours	12 months	200 hours	10 months	5 days
Data Transmission	USB, Bluetooth and ECHO	Bluetooth	Bluetooth	Bluetooth	ANT+	USB Bluetooth
Transmission Range (m)	300 meters	10 meters	10 meters	10 meters	3 meters	6 meters
Storage Capacity	480 hours	n/a	16 hours	65 hours	20 hours	30 days
Reported Variables	-Heart rate, -ECG, -Respiratory rate -Estimated core temperature -Activity levels -Posture -Peak acceleration	-Heart rate -Electrocardiogram-sensor	-Heart rate -Energy expenditure -Stride rate	-Heart rate	-Heart rate -Cadence	-Heart rate, -Steps taken, -Energy expenditure, -Distance, -Floors climbed, -Hourly activity, -Sleep,

	Fitbit Flex	Fitbit ONE	CT1	RT3	GT3X	GT1M	Jawbone UP3	Amiigo Band
Release Date	May 2013	Sep. 2012	Sep. 2005	Sep. 2005	May 2013	June 2011	Nov. 2014	Aug. 2013
Device Placement	Wrist	Waist	Waist	Waist	Waist	Wrist	Wrist	Wrist
Weight (g)	-	-	71.5	71.5	27	27	29	-
Operating Temperature	-20 – +45 °C	-20 – +45 °C	-	-	-	-	-	-
Accelerometer Type	Piezo-electric	Piezo-electric	Piezo-electric	Piezo-electric	n/a	n/a	-	-
Accelerometers axis	3	3	1	3	3	1	3	3
Battery Type	Lithium-polymer	Lithium-polymer	1.5V AAA	1.5V AAA	3.7 V Lithium	3.7 V Lithium	Lithium-polymer	Lithium-polymer
Battery Life	5 days	10-14 days	30 days	30 days	20 days	20 days	7 days	2 days
Data Transmission	USB & Bluetooth	USB & Bluetooth	USB	USB	USB	USB	USB & Bluetooth	USB & Bluetooth
Transmission Range (m)	6 meters	6 meters	-	-	-	-	6 meters	6 meters
Storage Capacity	30 days	23 days	21 days	21 days	40 days	40 days		30 days
Reported Variables	-Steps taken, -Energy expended -Floors climbed & Distance -Hourly activity, -Sleep,	Steps taken, -Energy expended -Floors climbed & Distance -Sleep,	-Energy expended	-Activity intensity, -METs,	-Step counts, -METs,	-Step counts, -Energy expended	-Resting heart rate and respiratory rates. -Step count, -Energy expenditure, -Sleep,	-Resting heart rate and respiratory rates. -Skin Temp. -Energy expended, -Sleep,

CHAPTER 2:

Title of Manuscript:

Psychometric Parameters of Zephyr Bioharness and Fitbit devices: A Systematic Review.

Authors:

Goris Nazari PT, MSc (C) [a],

Joy C MacDermid PT, PhD [b-c],

Kathryn E Sinden R. Kin., PhD [d],

Julie Richardson PT, PhD [a],

Ada Tang PT, PhD [a],

[a] McMaster University, School of Rehabilitation Science, 1400 Main Street West, Hamilton, Ontario, Canada.

[b] Physical Therapy, Western University, London, Ontario, Canada.

[c] Roth McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada

[d] Department of Kinesiology and Physical Education, McGill University, 475 Pine Street, Montreal, Québec, Canada.

Corresponding Author:

Goris Nazari, McMaster University, School of Rehabilitation Science, 1400 Main Street West IAHS 403, Hamilton, Ontario, Canada. E-mail: nazarigs@mcmaster.ca

Acknowledgements:

Joy C MacDermid was supported by a Canadian Institutes of Health Research (CIHR) Chair in Gender, Work and Health and the Dr James Roth Chair in Musculoskeletal Measurement and Knowledge Translation. This work was supported by an operating grant from the Ministry of Labour - Grant Number #13-R-027. Ada Tang was supported by a personnel award from the Heart and Stroke Foundation, Ontario Provincial Office (CS I 7468).

Ethics Approval:

This study was approved by Hamilton Integrated Research Ethics Board (HiREB) of McMaster University, Hamilton, Ontario, Canada**.**

Conflict of Interest:

There are no financial gains or conflict of interest with the content of this research article.

Abstract

Background: Easy and non-invasive methods to capture and export data on physiological status has become increasingly accessible due to innovations in wearable technologies. Wearable Physiological Monitoring devices have many applications for health and rehabilitation, but these benefits are dependent on reliability, validity and responsiveness of the data collected. Further, it is important to know that data from different devices can compared, if good levels of agreements are established.

Purpose: To synthesize and appraise the quality of published studies on psychometric parameters of Zephyr Bioharness and Fitbit devices.

Methods: We searched the Google Scholar and PubMed databases to identify articles. Articles were appraised for quality using a structured clinical measurement specific appraisal tool. Two raters evaluated the quality, and the primary author conducted the data extraction. We extracted data from the studies on the reliability (intra-class correlation coefficients and standard error of measurement) and validity measures (Pearson/Spearman's correlation coefficients) along with mean differences. To report responsiveness, Minimal Clinically Important Difference (MCID) and the cut-offs used to define importance was reported. Agreement parameters were summarised by the average biases and 95% limits of agreement.

Results: For Zephyr Bioharness, a total of ten studies were included; quality ranged from 42% to 79%. The intra-class correlation coefficients for Zephyr Bioharness heart rate and respiratory rates ranged from 0.97 – 0.98 and 0.30 – 0.90 respectively. The construct validity coefficients when compared against other standard measurement devices ranged

from 0.87 – 0.99 (heart rate) and 0.16 – 0.98 (respiratory rates). Agreement error of ≤ 2.40 and narrow 95% limits of agreement were reported for heart rate parameter when compared with gold standard measures. For Fitbit-ONE, a total of four studies were included; quality ranged from 32% to 82%. The intra-class correlation coefficients for Fitbit-ONE steps taken and energy expenditure ranged from 0.90 – 1.00 and 0.55 – 0.98 respectively. Construct validity coefficients ranged from 0.97 – 0.99 (steps taken) and 0.76 – 0.87 (energy expenditures) when comparing devices. An agreement bias of 155.90 between Fitbit-ONE energy expenditure and a gold standard metabolic system was reported.

Conclusion: Fair to very good quality evidence from ten studies suggests that the Zephyr Bioharness can provide a reliable and valid measurement of monitoring heart rate and other physiological measures across multiple contexts. Fair to very good evidence from four studies suggests that the Fitbit-ONE device demonstrated reliable and valid measures of steps taken and energy expenditures and displayed better agreements with gold standard measures than with the bronze.

Introduction

Study Rationale:

Wearable Physiological Monitoring (WPM) devices provide an easy and non-invasive method of capturing and transmission of wide range of physiological measures on the wearer (Johnstone et al., 2012). Hence, leading to further improvements in work place, athletic performance and overall quality of health (Gatti et al., 2014; Flanagan et al., 2014). Zephyr Bioharness ™ (Zephyr Technology Corporation, Annapolis, MD, US) and Fitbit (Fitbit Inc., San Francisco, CA, US) are among the most common WPM devices. The Zephyr Bioharness is capable of monitoring of physiological measures; heart rate, respiratory rate, activity levels, posture and estimated core temperature, whereas Fitbit captures variables such as; heart rate, number of steps taken, total energy expenditures, distance travelled and sleep time/efficiency (Zephyr Technology 2012; Fitbit Charge HR 2015).

It is important for a device to be reliable, valid and (if used to assess change over time) responsive, prior to its utilization (Streiner & Norman 1995). These clinical measurement properties are often referred to as the psychometric properties of a device. A device is said to be reliable if multiple measures under identical conditions yield similar results, whereas validity property of a device is when it measures what it was supposed to measure (Last 1983; Fletcher, Fletcher & Wagner 1988; Streiner & Norman 1995). Responsiveness of a device is referred to as the ability of a device to accurately detect change when it has occurred (De Bruin et al. 1997; Wright & Young 1998). In addition, it is necessary to establish the levels of agreement between devices to determine the extent to which the new

device varies from the "gold standard" measure, and whether the new device can replace the standard measure (Bland & Altman 1986). The data on psychometric parameters of WPM devices are accumulating. Therefore, a need exists to understand what is known about the psychometric parameters of WPM devices.

Study Purpose:

The aims of this systematic review were to synthesize published studies to describe the following with respect to the Zephyr Bioharness, Fitbit Charge Heart Rate and Fitbit-ONE devices 1) the volume, focus and quality of published studies, 2) the psychometric properties reported and 3) the gaps in the literature regarding their measurement properties.

Methods

Search

To identify articles on psychometric parameters of Zephyr Bioharness and Fitbit devices, we searched the Google Scholar and PubMed databases using the following keywords: Zephyr Bioharness reliability, validity, response, responsiveness and agreement, Fitbit Charge Heart Rate / Fitbit-ONE reliability, validity, responsiveness and agreement. Further articles were also identified by examining the reference list of each selected study. We were specifically interested in Zephyr Bioharness, Fitbit Charge Heart Rate and Fitbit-ONE devices, which have been introduced into the market between years 2010 – 2015. We chose Google scholar to gather articles that have been electronically published in the past six years because Google uses algorithms that can retrieve articles published in journals

regardless of the database they are indexed in (Al-Ubaydli 2005; Blakeman 2013), whereas a specific database is limited to the articles from the journals indexed in that specific database. We also searched PubMed as a standard source for health-related review.

Selection of Studies

At the first stage, two authors independently identified and screened Title/abstract. Studies that had used the devices to monitor physiological measures only, without reporting of psychometric parameters were considered irrelevant. An article was accepted if it met following specific eligibility criteria:

Inclusion Criteria:

1. Purpose of the study states assessing reliability or validity or responsiveness or agreement parameters, of Zephyr Bioharness or Fitbit Charge Heart Rate or Fitbit-ONE devices in healthy or clinical population.
2. Articles published in English,

Exclusion Criteria:

1. No data on the psychometric properties of Zephyr Bioharness or Fitbit Charge Heart Rate or Fitbit-ONE devices.
2. Studies that had used Zephyr Bioharness or Fitbit Charge Heart Rate or Fitbit-ONE devices to monitor physiological measures only.

Data Extraction

The primary author conducted the data extraction. For reliability measures, Standard Error of Measurement (SEM), intra-class correlation coefficient (ICC), mean differences and confidence intervals were extracted (MacDermid et al., 2009). These were interpreted using a common benchmark where ICC < 0.40 indicate poor, 0.40 ≤ ICC <0.75 indicate fair to good and ICC ≥ 0.75 indicate excellent reliability (Rosner 2005). For construct validity where these devices were compared against a reference standard, Pearson's/Spearman's correlation coefficients and mean difference data were extracted (MacDermid et al., 2009). The absolute value for the strength of the correlation were determined using the guide suggested by Evans (1996) as follows; 0.00-0.19 "very weak", 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong", 0.80-1.00 "very strong" (Evans 1996). To establish responsiveness, Minimal Important Difference (MID), or Minimal Clinically Important Difference (MCID), or Clinically Important Difference (CID) and the cut-off used to define importance was extracted (MacDermid et al., 2009). To assess levels of agreement, average agreement bias along with 95% Limits of Agreement (LoA) uniquely evaluate whether there is a discrepancy (bias) between two different devices measuring the same construct (Bunce 2009).

Quality Appraisal:

The articles were appraised by the first and second authors for quality using a structured clinical measurement specific appraisal tool (MacDermid et al., 2009). (Appendix D). The evaluation criteria included: 1) Thorough literature review to define the research question;

2) Specific inclusion/exclusion criteria; 3) Specific hypotheses; 4) Appropriate scope of psychometric properties; 5) Sample size; 6) Follow-up; 7) The authors referenced specific procedures for administration, scoring, and interpretation of procedures; 8) Measurement techniques were standardized; 9) Data were presented for each hypothesis; 10) Appropriate statistics-point estimates; 11) Appropriate statistical error estimates; and 12) Valid conclusions and clinical recommendations (MacDermid et al., 2009). An article's total quality score was calculated by summing of scores for each item, divided by the numbers of items and multiplied by 100% (MacDermid et al., 2008).

Results

A total of 592 studies were identified from the search in the databases [Google Scholar (n = 426) and PubMed (n = 166)], of which 56 studies were considered relevant. All 56 studies were retrieved and assessed for eligibility, and a total of 14, [Zephyr Bioahrness (n = 10), Fitbit-ONE (n = 4)] studies were included in this review (Figure 2.1). Table – 2.1 & 2.2 display the summary of the studies addressing the psychometrics of Zephyr Bioharness and Fitbit-ONE devices respectively. Quality of individual articles for Zephyr Bioharness was variable, ranging from 42% to 79% with 50% of articles reaching or exceeding a score of 67% on the quality rating (Table – 2.3). The most common flaws noted in the psychometric studies were 1) lack of sample size calculation/justification, and 2) not considering an appropriate scope of psychometric properties. In regards to Fitbit-ONE, the quality of individual studies ranged from 32% to 82% (Table – 2.4). Lack of reporting of specific

hypothesis and making general conclusion/clinical recommendations were observed as the most source of errors.

Reliability Parameters of Zephyr BioHarness:

We located three studies that examined the test-retest reliability measures of Zephyr Bioharness between years 2012 – 2015 [Table – 2.5] (Johnstone et al., 2012; Johnstone et al., 2012b; Rawstorn et al., 2015). The physical activities during which the reliability measures were studied included unstructured mobility; vacuuming and sweeping, and structure running/walking and cycling. The populations studied included young healthy recreational active males and females as well as older patients with atrial fibrillation.

Johnstone et al. 2012b reported the test-retest reliability of all five Zephyr Bioharness variables in ten physically active males, age 20.50 ± 2.10 years, weight 70.40 ± 9.40 kg and height 1.77 ± 0.10 m, during a treadmill running exercise protocol. The ICC (mean differences) of 0.98 (2.70 beats/minute), 0.75 (-0.52 breaths/minute), 0.99 (0.00 $ct \cdot sec^{-1}$), 0.61 (0.50 – degrees Celsius) and 0.99 (0.30 $^{\circ}$) for heart rate, respiratory rate, activity levels, temperature and posture variables were reported respectively.

Johnstone et al. (2012) reported the test-retest reliability of Zephyr Bioharness variables; heart rate, respiratory rate and activity levels, in ten physically active males, age 21.50 ± 2.80 years, weight 71.40 ± 7.90 kg and height 1.79 ± 0.10 m, during a Walk-Jog-Run exercise protocol. ICC (mean differences) of 0.97 (4.30 beats/minute), 0.90 (-0.51 breaths/minute) and 0.92 (-0.02 $ct \cdot sec^{-1}$) for heart rate, respiratory rate and activity level variables were reported.

Rawstorn et al. (2015), reported the reliability measures in two phases. The first phase tested 10 people who were recreationally active including 6-males and four females, age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m, with ICCs of 0.98 and 0.94 for heart rate and respiratory rates during a treadmill running test respectively. In the second phase of testing, five males and three females with atrial fibrillation, age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m were assessed during treadmill and cycle ergometer tests, and stimulated activities of daily living (sweeping and vacuuming). ICCs of 0.98 and 0.55 were reported for heart rate and respiratory rate variables respectively.

Overall, Zephyr Bioharness demonstrated high reliability scores, however inclusion of young males, small sample sizes and reporting of reliability measures only during activity, reflect upon the limited number of parameters evaluated in these studies (Johnstone et al., 2012; Johnstone et al., 2012b; Rawstorn et al., 2015).

Validity Parameters of Zephyr BioHarness:

We identified nine studies that assessed the validity of Zephyr Bioharness against criterion measures; gold or bronze standards, between years 2012 – 2015 [Table – 2.6] (Hailstone & Kilding 2011); Johnstone et al., 2012a; Johnstone et al., 2012; Kim et al., 2013; Gatti et al., 2014; Flanagan et al. 2014; Dolezal et al., 2014; Smith et al., 2014; Rawstorn et al., 2015). The phases during which validity measures were established included; rest/static (sitting), physical activities (unstructured mobility; vacuuming, sweeping, arm lifting, trunk rotation, crawl/search, and structure running/walking and cycling) and recovery. The populations

studied included healthy recreational active males and females as well as older patients with atrial fibrillation.

Flanagan et al. (2014) established the concurrent validity of Zephyr Bioharness variable; heart rate, in seventy-five healthy men, age 23.00 ± 4.00 years, height 1.81 ± 0.08 m and weight 83.00 ± 12.00 kg, against its gold standard; 5-Lead electrocardiogram (ECG), at rest, during a standardised 12-stage cycle ergometer exercise protocol and throughout recovery. Pearson correlation coefficients of ≥ 0.99 and mean differences of ≤ 2.00 beats/minute for all three phases between the Zephyr Bioharness and the 5-Lead ECG (gold standard criterion measure) were reported.

Kim et al. (2013) determined the validity of Zephyr Bioharness variables; heart rate and respiratory rate, in twelve healthy men, age 25.50 ± 4.10 years, height 180.10 ± 0.07 m and weight 78.80 ± 13.90 kg, against their gold standards; 12-Lead ECG and Gas Exchange (Indirect calorimeter), respectively. Participants performed a graded treadmill exercise protocol at different intensities determined by the percentage of each participant's maximum aerobic capacity. Spearman's rank correlation coefficients of 0.92 and 0.97 for the heart rate and respiratory rate variables were reported respectively. Furthermore, mean difference ranging from -3.30 – 3.20 beats/minute and -1.30 – 0.60 breaths/minute were recorded for heart rate and respiratory rate variables respectively.

Dolezal et al. (2014) reported the criterion validity of Zephyr Bioharness variables; heart rate, in ten healthy men, age 21.00 ± 1.00 years, height 1.84 ± 0.05 m and weight 91.00 ± 10.00 kg, against its gold standard: 12-Lead ECG. In this study participants performed four activities; treadmill walking, crawl & search, stair climbing and fast paced outdoor-

walking. Pearson correlation coefficients of 0.99 for all four activities and mean difference ranging from -0.13 – 0.04 beats/minute were found between the Zephyr and criterion measure.

Smith et al. (2014) established the validity of Zephyr Bioharness variable; heart rate and respiratory rate, in eleven healthy men age 20.00 ± 1.00 years, height 1.80 ± 0.07 m and weight 82.00 ± 10.20 kg, against gold standards; 3-Lead ECG and Gas Exchange (Indirect calorimeter) respectively. Three activities: treadmill walk, crawl and search, and ascending/descending stairs were performed. Pearson correlation coefficients ranging from 0.99 – 0.95 and 0.16 – 0.59, and mean differences -1.70 – 0.40 and -2.40 – 0.70 were reported for heart rate and respiratory rate respectively. Lower respiratory rate correlation coefficient of 0.16 – for search activity, could be due to improper fitting of Zephyr chest strap and possible displacement during extensive upper extremity movement. Since 'search' activity involved crawling down a 26-meter hallway into a room and performing a right-handed secondary search for 4 minutes at a self-determine pace.

Gatti et al. (2014) determined the validity of Zephyr Bioharness variables; heart rate and respiratory rate, in seven healthy males and three females, age 23.80 ± 2.90 years, height 1.79 ± 0.08 m and weight 75.50 ± 10.70 kg, against their gold standards: 5-Lead ECG and Gas Exchange (Indirect calorimeter), during five activities respectively. These activities include; Static (seated), Thoracic rotation, Arm lifting, Batting, Weight moving and Walking. Pearson correlation coefficients (mean differences) of 0.99 (-0.78), 0.98 (-0.77), 0.94 (0.22), 0.76 (-2.51), 0.78 (-4.81) and 0.74 (-1.68) for heart rates and 0.98 (-0.19), 0.84

(0.08), 0.91 (-1.16), 0.83 (-0.75), 0.54 (-2.42) and 0.91 (-2.57) for respiratory rates were reported.

Rawstorn et al. (2015) reported the validity of Zephyr Bioharness variables; heart rate and respiratory rate, in six recreationally active males and four females, age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m, against their gold standards; 12-Lead ECG and Gas Exchange (Indirect calorimeter), respectively. Participants performed a treadmill running protocol at different intensities determined by the percentage of each participant's maximum aerobic capacity. Spearman's rank correlation coefficients of 0.92 and 0.87, and median biases of -1.30 and -0.88 for heart rate and respiratory rate were reported respectively. In the second phase of testing, five males and three females with atrial fibrillation, age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m were assessed during treadmill and cycle ergometer tests, and stimulated activities of daily living (sweeping and vacuuming). Spearman's rank correlation coefficients of 0.97 and 0.43, and median differences of -1.45 and 2.16 for heart rate and respiratory rates were reported respectively. In this study, Zephyr respiratory rate correlation coefficients of 0.43 might have been confounded due to reasons such as; inclusion of upper extremity activities (improper fitting) and possibly inclusion of older individuals since impairment in mechanism of pulmonary function are common in this population regardless of any pathological conditions.

Johnstone et al. (2012) established the concurrent validity of Zephyr Bioharness variables; heart rate and respiratory rate, in ten physically active males, age 21.50 ± 2.80 years, weight 71.40 ± 7.90 kg and height 1.79 ± 0.10 m, against their criterion measures; Polar-T31 chest

strap and Portable Metalyser, respectively. The Walk-Jog-Run protocol was carried out by each participant at different velocities. Pearson's correlation coefficients of 0.98 and 0.82, and mean differences of 0.00 and -1.20 for heart rate and respiratory rate all velocities were reported respectively.

Johnstone et al. (2012a) determined the concurrent validity of Zephyr Bioharness variables; heart rate and respiratory rate, in twelve physically active males, age 21.50 ± 2.80 years, weight 71.40 ± 7.90 kg and height 1.79 ± 0.10 m, against their criterion measures; Polar-T31 chest strap and Gas Exchange (Indirect calorimeter), respectively. A treadmill running protocol was performed at different velocities by each of the study participants. Pearson's correlation coefficient of 0.89 and 0.91, and mean differences of -3.80 and -2.00 for heart rate and respiratory rate all velocities were reported respectively.

In summary, the device heart rate variable demonstrated very strong correlations at rest (Kim et al., 2013; Gatti et al., 2014; Flanagan et al., 2014), strong to very strong correlations during various activities; treadmill running, cycle ergometer, vacuuming, sweeping, thoracic rotation, arm lifting, batting, weight moving, walking, crawling, stair climbing, brisk walking (Kim et al., 2013; Gatti et al., 2014; Flanagan et al., 2014; Dolezal et al., 2014; Smith et al., 2014; Rawstorn et al., 2015) and again very strong correlations throughout recovery (Flanagan et al., 2014) when compared with adopted criterion measure; ECG. Similarly, Zephyr device respiratory rate variable displayed very strong correlations at rest (Kim et al., 2013; Gatti et al., 2014) and during wide range of activities when examined against adopted criterion measures; Gas Exchange (Indirect calorimeter),

(Johnstone et al., 2012; Johnstone et al., 2012a; Kim et al., 2013; Gatti et al. 2014; Rawstorn et al., 2015).

Responsiveness Parameters of Zephyr BioHarness:

No studies were identified.

Agreement Parameters of Zephyr BioHarness:

We identified eight studies that examined the pair-wise comparison of Zephyr Bioharness with adopted criterion measures; gold or bronze standards, between years 2012 – 2015 [Table – 2.7] (Johnstone et al., 2012; Johnstone et al., 2012a; Kim et al., 2013; Gatti et al., 2014; Dolezal et al., 2014; Smith et al., 2014; Flanagan et al., 2014; Rawstorn et al., 2015). Four studies reported heart rate and respiratory rate agreement biases of ≤ 2.40 and narrow 95% limits of agreement in pairwise device comparison of Zephyr at rest and recovery phases and during various activities including: thoracic rotation, arm lifting, treadmill walking, cycle ergometer, crawling, stair climbing (Gatti et al., 2014; Flanagan et al., 2014; Dolezal et al., 2014; Smith et al. 2014).

Two studies with similar pair-wise device comparisons for heart rate and respiratory rates, reported agreement biases of ≤ 1.25 with wider 95% limits of agreement during treadmill running/walking, cycle ergometer, batting, weight moving, sweeping and vacuuming activities (Rawstorn et al., 2015; Kim et al., 2013)

The inter-device agreement was examined between Zephyr and Polar T31 (bronze standard) heart rate measures. Agreement biases of ≤ 3.05 with much wider 95% limits of

agreement were reported during a treadmill walk/run testing protocol (Johnstone et al., 2012; Johnstone et al., 2012a).

Overall, Zephyr Bioharness displayed better agreements with gold standard measures than with the bronze, therefore, suggestive of possible interchangeable use.

Reliability Parameters of Fitbit-ONE:

We found three studies that assessed the test-retest reliability measures of Fitbit-ONE [Table – 2.8] (Takacs et al., 2014; Diaz et al., 2015; Ferguson et al., 2015).

Diaz et al. (2015) reported the test-retest reliability of Fitbit-ONE variables; steps taken and energy expenditure, in twenty-three participants (13 females), ranging from 20 – 54 years and body mass index range 19.60-29.90 kg/m^2. Participants performed a treadmill-walk/jog exercise protocol at various speeds. Overall, test-retest reliability measures – ICCs of 0.99 and 0.96 for the step count and energy expenditure variables were established respectively.

Takacs et al. (2013) determined the test-retest reliability Fitbit-One – steps taken variable, in thirty participants (15 females), age 29.60 ± 5.70 years and body mass index 22.70 ± 3.00 kg/m^2. All participants performed a treadmill exercise protocol at various speeds 0.90 m/s – 1.78 m/s and ICC of ≥ 0.95 and mean difference of ≤ 3.00 (steps) were reported.

Fergusson et al. (2015) reported the test retest reliability of Fitbit-ONE variables; step taken and energy expenditure, among twenty-one healthy participants (11 females), age 32.80 ±10.20 years and body mass index 25.50 ± 5.20 kg/m^2, during a 48-hour period involving

activities of daily living. ICCs (mean differences) of 0.95 (779.00) and 0.55 (349.00) were reported for steps taken and energy expenditure variables respectively.

To summarise, Fitbit-One device displayed excellent test-retest reliability measures for steps taken and good to excellent measures for energy expenditures variables during different exercise protocols (Takacs et al., 2014; Diaz et al., 2015; Ferguson et al., 2015).

Validity Parameters of Fitbit-ONE:

We identified four studies that examined the validity of Fitbit-ONE device [Table – 2.9] (Lee et al., 2014; Takacs et al., 2014; Diaz et al., 2015; Ferguson et al., 2015).

Lee et al. (2014) established the criterion validity of Fitbit-One – energy expenditure variable, in sixty healthy participants (30 females), age 26.00 ± 5.70 years, against its criterion measure; Portable metabolic system. Study participants performed a 69-minute exercise protocol that included; sedentary, walking, running and moderate to vigorous activities. An overall Pearson correlation coefficient of 0.81 and mean differences of 26.00 (calories), were reported between Fitbit-One and its adopted criterion measure.

Diaz et al. (2015) reported the validity of Fitbit-One variables; steps taken and energy expenditure, in twenty-three participants (13 females), ranging from 20 – 54 years and body mass index range 19.60-29.90 kg/m^2, against their criterion measures; Direct observational count (of number of steps) and Gas Exchange (Indirect calorimeter) respectively. During a treadmill-walk/jog exercise protocol, Pearson correlation coefficients of 0.97 - 0.99 and 0.86 - 0.87 for the step count and energy expenditure variables were reported respectively.

Fergusson et al. (2015) reported the convergent validity of Fitbit-One variables; step taken and energy expenditure, during a 48-hour period involving activities of daily living in twenty-one healthy participants (11 females), age 32.80 ±10.20 years and body mass index 25.50 ± 5.20 kg/m^2, against their bronze standard criterion measures; Acti-Graph GT3X and Body Media Sense-Wear Model, respectively. Pearson correlation coefficients of 0.99 and 0.76 for the steps taken and energy expenditure variables were recoded respectively.

Takacs et al. (2013) determined the validity of Fitbit-One – steps taken variable, in thirty participants (15 females), age 29.60 ± 5.70 years and body mass index 22.70 ± 3.00 kg/m^2, against its gold standard criterion measure; Direct observational count. All participants performed a treadmill exercise protocol at various speeds and Pearson correlation coefficients of ≥ 0.97 and mean difference of 0.40 (steps) were established.

Overall, Fitbit-ONE device – steps taken and energy expenditure variables, demonstrated very strong correlations with its adopted criterion measures during various activities (Lee et al., 2014; Takacs et al., 2014; Diaz et al., 2015; Ferguson et al., 2015).

Responsiveness Parameters of Fitbit-ONE:

No studies were identified.

Agreement Parameters of Fitbit-ONE:

We identified two studies that examined the pair-wise comparison of Fitbit-ONE with adopted criterion measures; gold and bronze standards, between years 2014 – 2015 [Table – 2.10] (Lee et al., 2014; Ferguson et al., 2015).

Lee et al. (2014) study reported agreement mean bias of 155.90 cal. in pairwise device comparison of Fitbit-ONE energy variable with gold standard measure – Metabolic system, in sixty healthy participants (30 females), age 26.00 ± 5.70 years, during a 69-minute exercise protocol that included; sedentary, walking, running and moderate to vigorous activities.

Fergusson et al. (2015) reported the pairwise device comparison of Fitbit-ONE variables; step taken and energy expenditure, against their bronze standard criterion measures; Acti-Graph GT3X and Body Media Sense-Wear Model, respectively. Mean agreement biases of 584.0 steps and 475.0 Kcal. and wide 95% limits of agreement were established among twenty-one healthy participants (11 females), age 32.80 ±10.20 years and body mass index 25.50 ± 5.20 kg/m^2 during a 48-hour period involving activities of daily.

Overall, Fitbit-ONE displayed better agreements with gold standard measures than with the bronze standard measures (Lee et al., 2014; Ferguson et al., 2015).

Psychometric Parameters of Fitbit Charge

No studies were identified.

Discussion

This study synthesized current research in 14 studies addressing the psychometric parameters of the Zephyr Bioharness and Fitbit-ONE devices and was able to provide some clinical recommendation regarding its use. This systematic review identified fair to very good quality evidence for suggesting that the Zephyr Bioharness and Fitbit-ONE can

provide a reliable and valid measurement of monitoring heart rate and other physiological measures across multiple contexts. Overall, although this review identified substantial gaps in the literature, the emerging studies from a limited number of applications are generally supportive of using Zephyr or Fitbit-ONE devices in clinical research applications and field-based monitoring of physiological and activity parameters is required. Wearable technologies that are cost accessible for individuals to use in health promotion or therapeutic applications has opened up a wide range of applications that are relevant to physical therapy in both primary and secondary prevention of disability. Devices that can monitor physiological responses might be important to allow physical therapist to more precisely prescribed exercise and determine thresholds for progression. These devices also allow patients more independence in obtaining physiological or activity targets; and may promote better adherence and treatment fidelity. While much research will be needed to fully define the patient populations, applications and impact on outcomes that can be obtained by using these wearable technologies, it is clear that this is an issue of importance to physical therapy.

Based upon this systematic review, we have identified and classified the gaps in the current literature. There is a need for future studies that would address measurement properties of the devices with respect to the following:

1. during different types of activities or clinical tests,
2. in different populations; healthy/patient, older/younger, males/females.
3. in different physiological states including resting, recovery, low level or vigorous exercise, and activities of daily living.

In the following sections, we have reported the specific literature gaps concerning the Zephyr Bioharness, Fitbit-ONE and Fitbit Charge Heart Rate devices.

Zephyr Bioharness Device:

1. The responsiveness of Zephyr Bioharness.
2. The reliability and responsiveness of Zephyr Bioharness in healthy male and female participants across various age groups.
3. The reliability, validity, responsiveness and agreement properties of Zephyr Bioharness in clinical populations.
4. The reliability and responsiveness of Zephyr Bioharness in resting, activity and recovery states.

Fitbit-ONE Device:

1. The responsiveness of Fitbit-ONE.
2. The reliability, validity, responsiveness and agreement properties of Fitbit-ONE in clinical populations.
3. The responsiveness and agreement parameters of Fitbit-ONE in resting, activity and recovery states.

Fitbit Charge Heart Rate Device:

1. The reliability, validity, responsiveness and agreement properties of Fitbit Charge Heart Rate device.
2. The reliability, validity, responsiveness and agreement properties of Fitbit Charge Heart Rate device in normal male and female participants across various age groups.
3. The reliability, validity, responsiveness and agreement properties of Fitbit Charge Heart Rate device in patient population.
4. The reliability, validity, responsiveness and agreement properties of Fitbit Charge Heart Rate device in resting, activity and recovery states.

In this thesis project, we intend to address several gaps that exist in the current literature.

1. To establish the reliability measures of Zephyr Bioharness and Fitbit Charge Heart Rate devices in a large sample of healthy male and female participants during three states; rest, submaximal activity and recovery.
2. To determine the validity and agreement parameters between Zephyr Bioharness and Fitbit Charge Heart Rate devices in a large sample of healthy male and female participants during three states; rest, submaximal activity and recovery.

We used Google scholar database to identify articles on psychometric properties and agreement parameters of these devices. We chose Google scholar over other databases mainly because it uses algorithms that can identify all electronically published articles from various journals and not just articles from journals that are indexed in a specific database, which is the case with most databases (Al-Ubaydli 2005; Blakeman 2013).

Conclusion

Fair to very good quality evidence from ten studies suggests that the Zephyr Bioharness can provide a reliable and valid measurement of monitoring heart rate and other physiological measures across multiple contexts. Fair to very good evidence from four studies suggests that the Fitbit-ONE device demonstrated reliable and valid measures of steps taken and energy expenditures and displayed better agreements with gold standard measures than with the bronze standard measures in healthy population.

References:

Al-Ubaydli, M., (2005). Using Search Engines to Find Online Medical Information. PLoS medicine, 2, 842-844.

Blakeman, K., (2013). Finding research information on the web: how to make the most of Google and other free search tools. Science Progress, 96, 61-84.

Bland, J.M., & Altman, D. G., (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **327** (8476), 307–10.

Bunce, C., (2009). Correlation, Agreement, and Bland–Altman Analysis: Statistical Analysis of Method Comparison Studies. *Am J Ophthalmol*,148(1), 4-6.

De Bruin, A. F., Diederiks, J.P.M., De Witte, L.P., Stevens, F.C.J., Philipsen, H., (1997). Assessing the responsiveness of a functional status measure: The Sickness Impact Profile versus the SIP68. J Clin Epidemiol, 50(5):529–40.

Diaz, K.M., Krupka, D.J., Chang, M.J., Peacock, J., Ma, Y., Goldsmith, J., … Davidson, K.W., (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, 185, 138–140.

Dolezal, B.A., Boland, D.M., Carney, J., Abrazado, M., Smith, D.L. and Cooper, C.B., (2014). Validation of heart rate derived from a physiological status monitor-embedded compression shirt against criterion ECG. *Journal of Occupational and Environmental Hygiene*, 11: 833–839.

Evans, J.D. (1996). *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Ferguson, T., Rowlands, A. V., Olds, T. and Maher, C., (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity,* 12 (42), 201-209.

Fitbit charge HR Wireless Heart Rate + Activity Wrist Band (2015, January*). Fitbit charge HR$_{TM}$ Product Manual Version 1.0.* Retrieved from https://www.fitbit.com/ca/chargehr.

Flanagan, S.D., Comstock, B.A., Dupont, W.H., Sterczala, A.R., Looney, D.P., Dombrowski, D.H., … Kraemer, W.J. (2014). Concurrent validity of the armour39 heart rate monitor strap. *Journal of Strength and Conditioning Research,* 28(3), 870–873.

Fletcher, R.H., Fletcher, S.W., Wagner, E.H. (1988). Clinical Epidemiology—The Essentials, 2nd ed. Baltimore: Williams & Wilkins.

Gatti, U.C., Schneider, S. and Migliaccio, G.C. (2014). Physiological condition monitoring of construction workers. *Automation in Construction* 44, 227–233.

Hailstone, J, & Kilding, A.E. (2011). Reliability and Validity of the Zephyr™ BioHarness™ to Measure Respiratory Responses to Exercise. *Measurement in Physical Education and Exercise Science*, 15: 293–300.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012a). Bioharness™ multivariable monitoring device: Part I: Validity. *Journal of Sports Science and Medicine* 11(3), 400- 408.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T., Mitchell A.C. and Garrett, A.T. (2012). Field based reliability and validity of the Bioharness™ multivariable monitoring device. *Journal of Sports Science and Medicine* 11, 643-652.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012b). Bioharness™ multivariable monitoring device: Part II: Reliability. *Journal of Sports Science and Medicine* 11(3), 409- 417.

Kim, J.H., Roberge, R., Powell, J. B., Shafer, A. B. and Williams, W. J. (2013). Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr Bioharness ™. *Int J Sports Med,* 34, 497–501.

Last, J. M. (1983). A Dictionary of Epidemiology. New York: Oxford University Press.

Lee, J. M., Kim, Y., and Welk, G. J., (2014). Validity of consumer-based physical activity monitors. *Journal of the American College of Sports Medicine,* 14, 1840-1848.

MacDermid, J.C., Walton, D.M., Avery, S., Blanchard, A., Etruw, E., Mcalpine, C., Goldsmith, C. H., (2009). Measurement Properties of the Neck Disability Index: A Systematic Review. *Journal of orthopaedic & sports physical therapy*, 39 (5), 400 – 416.

Rawstorn, J.C., Gant, N., Warren, I., Doughty, R.N., Lever, N., Poppe, K.K. and Maddison, R. (2015). Measurement and data transmission: Validity of a multi-biosensor system for real-time remote exercise monitoring among cardiac patients. *JMIR Rehabil Assist Technol,* 2 (1): e2.

Rosner, B. (2005). Fundamentals of biostatistics (6th ed.). Boston: Duxbury Press.

Smith, D.L., Haller, J.M., Dolezal, B.A., Cooper, C.B. and Fehling, P.C. (2014). Evaluation of a wearable physiological status monitor during simulated fire fighting activities. *Journal of Occupational and Environmental Hygiene*, 11: 427–433.

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: a practical guide to their development and use* (2nded). New York: Oxford University Press.

Takacs, J., Pollock, C.L., Guenther, J.R., Bahar, M., Napier, C. and Hunt, M.A., (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17, 496–500

Wright, J.G, and Young, N.L., (1998). A comparison of different indices of responsiveness. *J Clin Epidemiol*, 50(3):239–46.

Zephyr Technology (2012, September 12). *BioHarness 3.0 User ManualBioHarness 3.0 User Manual*. Retrieved from http://zephyranywhere.com/products/bioharness-3/

Table – 2.1 Summary of Studies Addressing Psychometrics of Zephyr Bioharness Device

Authors	Sample	n	Properties Evaluated
Rawstorn et al. (2015) Phase (I)	Six recreationally active males and four females, Age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m.	10	Reliability, validity & agreement.
Rawstorn et al. (2015) Phase (II)	Five males and three females with atrial fibrillation, Age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m.	8	
Johnstone et al. (2012)	Ten physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.1 m	10	Reliability, validity (concurrent) & agreement.
Johnstone et al. 2012b	Ten physically active males, age 20.5 ± 2.1 years, weight 70.4 ± 9.4kg and height 1.77 ± 0.10 m	10	Reliability
Gatti et al. (2014)	Seven healthy males and three females, Age 23.8 ± 2.9 years, height 179 ± 8 cm and weight 75.5 ± 10.7 kg.	10	Validity & agreement
Smith et al. (2014)	Eleven healthy men age 20 ± 1 years, height 1.80 ± 0.07 m and weight 82.0 ± 10.2 kg	11	Validity & agreement
Dolezal et al. (2014)	Ten healthy men, age 21 ± 1 years, height 184 ± 5 cm and weight 91 ± 10 kg.	10	Validity (criterion) & agreement
Flanagan et al. (2014)	Seventy-five healthy men, age 23 ±4 years, height 181 ±8 cm and weight 83±12 kg	75	Validity (concurrent) & agreement
Kim et al. (2013)	Twelve healthy men, age 25.5±4.1, height 180.1±6.5 and weight 78.8 ±13.9	12	Validity & agreement
Johnstone et al. (2012a)	Twenty-two physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.10 m	22	Validity (concurrent) & agreement
Hailstone & Kilding (2011).	Six recreationally active males and six females, 20.0 ± 1.2 years, 75.3 ± 11.5 weight and 176.5 ± 10.1 height.	12	Reliability & validity

Table – 2.2 Summary of Studies Addressing Psychometrics of Fitbit-ONE Device.

Authors	Sample	n	Properties Evaluated
Diaz et al. (2015)	Twenty-three healthy adult participants, (10 males, 13 females), age 20 – 54 years, body mass index 19.60 – 29.90 kg/m^2	23	Reliability & validity
Takacs et al. (2013)	Thirty subjects (15 males. 15 females). Age 29.60 ±5.70 years and body mass index 22.70 ± 3.00 kg/m^2	30	Reliability & validity
Ferguson et al. (2015)	Twenty-One participants, 10 males (BMI 27.30 ± 3.20), 11 females BMI 25.50 ± 5.20, mean age 32.80 ± 10.20 years	21	Reliability, validity(convergent) & agreement
Lee et al. (2014)	Sixty healthy participants, (Males 30, Age 28.6±6.40 years, Height 176.10 ± 5.40cm, Weight 75.40 ±9.50 kg, body mass index 24.30±2.60 kg/m^2). (Females 30, Age 24.2±4.70 years, Height 166.00 ± 7.00cm, Weight 60.30 ±8.60 kg, body mass index 21.80±2.70 kg/m^2)	60	Validity (criterion) & agreement

Table – 2.3 Quality of Studies on The Psychometric of Zephyr Bioharness Device

Study	Item Evaluation Criteria*												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total (%)**
Dolezal et al. 2014	1	1	1	0	0	N/A	2	2	2	2	2	2	68 %
Smith et al. 2014	1	1	0	0	0	N/A	2	2	2	2	2	1	59 %
Rawstorn et al. 2015	1	1	0	2	0	0	2	2	2	2	2	2	67 %
Flanagan et al. 2014	1	1	1	0	1	N/A	2	2	1	2	2	1	64 %
Gatti et al. 2014	1	2	0	0	0	N/A	2	2	2	2	2	1	64 %
Kim et al. 2013	2	1	0	0	0	N/A	2	2	2	2	2	1	64 %
Hailstone & Kilding 2011	0	1	0	1	0	0	2	2	2	1	0	1	42 %
Johnstone et al. 2012	2	1	1	2	0	2	2	2	2	2	2	1	79 %
Johnstone et al. 2012a	2	1	1	0	0	N/A	2	2	2	2	2	1	68 %
Johnstone et al. 2012b	2	1	1	0	0	2	2	2	2	2	2	1	71 %

**Item Evaluation Criteria: 1. Thorough literature review to define the research question; 2. Specific inclusion/exclusion criteria; 3. Specific hypotheses; 4. Appropriate scope of psychometric properties; 5. Sample size; 6. Follow-up; 7. The authors referenced specific procedures for administration, scoring, and interpretation of procedures; 8. Measurement techniques were standardized; 9. Data were presented for each hypothesis; 10. Appropriate statistics-point estimates; 11. Appropriate statistical error estimates; 12. Valid conclusions and clinical recommendations.*

Table – 2.4 Quality of Studies on The Psychometric of Fitbit-ONE Device.

Study	Item Evaluation Criteria*												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total (%)**
Diaz et al. 2015	1	1	0	1	0	N/A	1	1	0	1	1	0	32 %
Takacs et al. 2014	1	2	0	1	2	N/A	2	2	1	2	2	1	73 %
Ferguson et al. 2015	2	2	2	0	2	N/A	2	2	2	2	1	1	82 %
Lee et al. 2014	1	1	2	0	0	N/A	2	2	2	2	1	1	64 %

**Item Evaluation Criteria: 1. Thorough literature review to define the research question; 2. Specific inclusion/exclusion criteria; 3. Specific hypotheses; 4. Appropriate scope of psychometric properties; 5. Sample size; 6. Follow-up; 7. The authors referenced specific procedures for administration, scoring, and interpretation of procedures; 8. Measurement techniques were standardized; 9. Data were presented for each hypothesis; 10. Appropriate statistics-point estimates; 11. Appropriate statistical error estimates; 12. Valid conclusions and clinical recommendations.*

Table – 2.5 Reliability measures of Zephyr Bioharness.

Authors	Zephyr Variables	Testing Protocol	ICC	SEM	C. I.	Mean Diff.	Sample Demographic
Rawstorn et al. (2015) Phase (I)	-Heart Rate (beats/min).	Treadmill Running	0.98	5.20	-	-	6 recreationally active males and 4 females, Age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m.
	-Breathing Rate (br/min).		0.94	2.78	-	-	
Rawstorn et al. (2015) Phase (II)	-Heart Rate (beats/min).	Treadmill, Cycle Ergometer and Activities of daily living; sweep and vacuum	0.98	4.77	-	-	5 males and 3 females with atrial fibrillation, Age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m.
	-Breathing Rate (br/min).		0.55	4.60	-	-	
Johnstone et al. (2012)		Walk-Jog-Run	0.97	4.60	-4.56 – 3.92	4.30	Ten physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.1 m
	-Heart Rate (beats/min).	All Velocities	0.89	5.90	-2.40 – -1.23	-0.20	
		4 – 6 km/h	0.93	4.10	-5.55 – -4.71	5.10	
		8 – 10.5 km/h	0.85	2.80	-6.32 – -4.82	5.60	
		11.0 km/h					
			0.90	7.70	0.39 – 0.64	-0.50	
	-Breathing rate. (br/min).	All Velocities	0.65	10.10	0.81 – 1.37	-1.10	
		4 – 6 km/h	0.91	6.60	0.41 – 0.69	-0.50	
		8 – 10.5 km/h	0.30	7.30	-1.11 – -0.23	0.70	
		11.0 km/h					
			0.92	14.70	0.02 – 0.03	-0.02	
	-Activity Levels. ($ct \cdot sec^{-1}$).	All Velocities	0.53	14.50	0.01 – 0.03	-0.03	
		4 – 6 km/h	0.92	14.70	0.02 – 0.03	-0.02	
		8 – 10.5 km/h	0.53	14.50	0.01 – 0.03	-0.03	
		11.0 km/h					
Johnstone et al. 2012b	-Heart Rate. (beats/min)	Treadmill Running	0.98	4.80	-3.15 – 2.22	2.70	Ten physically active males, age 20.5 ± 2.1 years, weight 70.4 ± 9.4kg and height 1.77 ± 0.10 m
	-Breathing Rate. (br/min)		0.75	17.50	0.21 – 0.82	-0.50	
	-Activity Levels. ($ct \cdot sec^{-1}$).		0.99	6.50	-0.001 – 0.01	0.00	
	-Temperature (°C).	Tilt Table	0.61	3.70	-0.61 – 0.42	0.50	
	-Posture (°).		0.99	7.60	-0.31 – 1.00	0.30	

Table – 2.6 Validity measures of Zephyr Bioharness.

Authors	Zephyr Variables	Validity Criterion Measure	Testing Protocol	r / rho	Mean Differences	Sample Demographic
Rawstorn et al. (2015) Phase (I)	-Heart Rate (beats/min).	-12 Lead ECG	Treadmill Running	0.92	-1.30	6 recreationally active males and 4 females, Age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m.
	-Breathing Rate (br/min).	-Indirect calorimeter		0.87	-0.88	
Rawstorn et al. (2015) Phase (I)	-Heart Rate (beats/min).	-12 Lead ECG	Treadmill, Cycle Ergometer and Activities of daily living; sweep and vacuum	0.97	-1.45	5 males and 3 females with atrial fibrillation, Age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m.
	-Breathing Rate (br/min).	-Indirect calorimeter		0.43	2.16	
Gatti et al. (2014)	-Heart Rate (beats/min).	-5 Lead EKG	Static.	0.99	-0.78	7 healthy males and 3 females, Age 23.8 ± 2.9 years, height 179 ± 8 cm and weight 75.5 ± 10.7 kg.
			Thoracic rotation.	0.98	-0.77	
			Arm Lifting.	0.94	0.22	
			Batting.	0.76	-2.51	
			Weight moving.	0.78	-4.81	
			Walking.	0.74	-1.68	
	-Breathing Rate (br/min).	-Metabolic Cart	Static.	0.98	-0.19	
			Thoracic rotation.	0.84	0.08	
			Arm Lifting.	0.91	-1.16	
			Batting.	0.83	-0.75	
			Weight moving.	0.54	-2.24	
			Walking.	0.91	-2.57	
Smith et al. (2014)	-Heart Rate (beats/min).	-3 Lead ECG	Treadmill Walk	0.99	-0.40	Eleven healthy men age 20 ± 1 years, height 1.80 ± 0.07 m and weight 82.0 ± 10.2 kg
			Search (Crawl)	0.95	-1.70	
			Stairs	0.99	0.40	
	-Breathing Rate (br/min).	-Metabolic Measurement System	Treadmill Walk	0.59	0.70	
			Search (Crawl)	0.16	-1.80	
			Stairs	0.55	-2.40	

Table – 2.6 Validity measures of Zephyr Bioharness.

Authors	Zephyr Variables	Validity Criterion Measure	Testing Protocol	r / rho	Mean Differences	Sample Demographic
Dolezal et al. (2014)	-Heart Rate (beats/min).	-12 Lead ECG	Treadmill Walk Search (Crawl) Stairs Fast walk	0.99 0.99 0.99 0.99	0.04 -0.01 -0.13 0.03	Ten healthy men, age 21 ± 1 years, height 184 ± 5 cm and weight 91 ± 10 kg.
Flanagan et al. (2014)	-Heart Rate (beats/min).	-5 Lead ECG	Cycle ergometer Rest 12- Stages Recovery	 ≥0.99 ≥0.99 ≥0.99	 1.00 0.00 2.00	Seventy-five healthy men, age 23 ±4 years, height 181 ±8 cm and weight 83±12 kg
Kim et al. (2013)	-Heart Rate (beats/min).	-12-Lead ECG	Treadmill running Baseline 30 % VO_{2max} 50 % VO_{2max} 70% VO_{2max} 90% VO_{2max} VO_{2max}	 0.87 0.88 0.92 0.90 0.96 0.92	 3.20 4.4 -1.40 -3.30 0.60 -0.80	Twelve healthy men, age 25.5±4.1, height 180.1±6.5 and weight 78.8 ±13.9
	-Breathing Rate (br/min).	-Metabolic Cart	Baseline 30 % VO_{2max} 50 % VO_{2max} 70% VO_{2max} 90% VO_{2max} VO_{2max}	0.80 0.50 0.95 0.90 0.99 0.97	0.60 -1.20 -1.30 0.20 -0.30 -0.50	

Table – 2.6 Validity measures of Zephyr Bioharness.

Authors	Zephyr Variables	Validity Criterion Measure	Testing Protocol	r / rho	Mean Differences	Sample Demographic
Johnstone et al. (2012)	-Heart Rate (beats/min).	Polar-T31.	Walk-Jog-Run All Velocities 4 – 6 km/h 8 – 10.5 km/h 11.0 km/h	 0.98 0.92 0.93 0.67	 0.00 1.30 -0.70 -2.10	Ten physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.1 m
	-Breathing rate (br/min).	-Portable Metalyser.	All Velocities 4 – 6 km/h 8 – 10.5 km/h 11.0 km/h	0.82 0.60 0.70 0.83	-1.20 -0.60 -1.80 -1.50	
	-Activity Levels ($ct \cdot sec^{-1}$).	-Oxygen Expenditure.	All Velocities	0.91	-	
Johnstone et al. (2012a)	-Heart Rate (beats/min).	-Polar-T31.		0.89	-3.80	Twenty-two physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.10 m
	-Breathing rate (br/min).	-Portable Metalyser.		0.91	-2.00	
	-Activity Levels ($ct \cdot sec^{-1}$).	-Oxygen Expenditure.	Treadmill – Walking and Running	0.97	-	
	-Temperature (°C).	-Separate skin thermistor.	Cycle Ergometer	0.76	-0.20	
	-Posture (°).	-Leighton Flexometer	Inclinometer (tilt Table)	0.99	0.00	
Hailstone & Kilding (2011).	-Breathing rate (br/min).	-Metamax 3b system	Incremental Treadmill Exercise	0.96	2.20	Six recreationally active males and six females, 20.0 ± 1.2 years, 75.3 ± 11.5 weight and 176.5 ± 10.1 height.

Table – 2.7 Agreement parameters of Zephyr Bioharness.

Authors	Zephyr Variables	Criterion Measure	Testing Protocol	Agreement Bias	95% LoA	Sample Demographic
Rawstorn et al. (2015) Phase (I)	-Heart Rate (beats/min).	-12 Lead ECG -Indirect calorimeter	Treadmill Running	-0.30	-21.87 – 9.26	6 recreationally active males and 4 females, Age 26.68 ± 3.26 years, weight 71.10 ±11.53 kg and height 1.73 ± 0.06 m.
	-Breathing Rate (br/min).			-1.25	-13.73 – 9.41	
Rawstorn et al. (2015) Phase (I)	-Heart Rate (beats/min).	-12 Lead ECG -Indirect calorimeter	Treadmill, Cycle Ergometer and Activities of daily living; sweep and vacuum	1.10	-13.39 – 23.79	5 males and 3 females with atrial fibrillation, Age 69.68 ±9.53 years, weight 77.46 ±18.81 kg and height 1.69 ± 0.12 m.
	-Breathing Rate (br/min).			0.39	-11.58 – 18.91	
Gatti et al. (2014)	-Heart Rate (beats/min).	- 5 Lead EKG	Static.	-0.78	-5.10 – 3.60	7 healthy males and 3 females, Age 23.8 ± 2.9 years, height 1.79 ± 8 m and weight 75.5 ± 10.7 kg.
			Thoracic rotation.	-0.77	-5.40 – 3.90	
			Arm Lifting.	0.22	-7.70 – 8.10	
			Batting.	-2.51	-15.70 – 10.70	
			Weight moving.	-4.81	-22.50 – 12.90	
			Walking.	-1.68	-15.60 – 12.20	
	-Breathing Rate (br/min).	-Metabolic Cart	Static.	-0.19	-2.58 – 2.20	
			Thoracic rotation.	0.08	-6.69 – 6.86	
			Arm Lifting.	-1.16	-7.06 – 4.73	
			Batting.	-0.75	-5.95 – 4.46	
			Weight moving.	-2.24	-10.42 – 5.57	
			Walking.	-2.57	-15.10 – 9.97	
Smith et al. (2014)	-Heart Rate (beats/min).	- 3 Lead ECG	Treadmill Walk	-0.40	-0.70 – -0.10	Eleven healthy men age 20 ± 1 years, height 1.80 ± 0.07 m and weight 82.0 ± 10.2 kg
			Search (Crawl)	-1.70	-3.1 – -0.40	
			Stairs	0.40	0.04 – 0.70	
	-Breathing Rate (br/min).	-Metabolic Measurement System	Treadmill Walk	0.70	-0.10 – 1.50	
			Search (Crawl)	-1.80	-4.10 – 0.50	
			Stairs	-2.40	-3.40 – -1.40	

Table – 2.7 Agreement parameters of Zephyr Bioharness.

Authors	Zephyr Variables	Criterion Measure	Testing Protocol	Agreement Bias	95% LoA	Sample Demographic
Dolezal et al. (2014)	-Heart Rate (beats/min).	-12 Lead ECG	Treadmill Walk Search (Crawl) Stairs Fast walk (Outdoor)	0.04 -0.01 -0.13 0.03	-0.05 – 0.12 -0.12 – 0.10 -0.21 – -0.04 -0.09 – 0.14	Ten healthy men, age 21 ± 1 years, height 184 ± 5 cm and weight 91 ± 10 kg.
Flanagan et al. (2014)	-Heart Rate (beats/min).	-5 Lead ECG	Cycle ergometer (Rest, 12-stages of Cycle Ergometer and Recovery)	1.50	-2.84 – 2.42	Seventy-five healthy men, age 23 ±4 years, height 181 ±8 cm and weight 83±12 kg
Kim et al. (2013)	-Heart Rate (beats/min). -Breathing Rate (br/min).	-12-Lead ECG -Metabolic Cart	Treadmill Running	0.50 -0.60	-15.30 – 16.30 -5.60 – 4.40	Twelve healthy men, age 25.5±4.1, height 180.1±6.5 and weight 78.8 ±13.9
Johnstone et al. (2012a)	-Heart Rate (beats/min). -Breathing rate (br/min). -Temperature (°C) -Posture (°)	-Polar-T31. -Portable Metalyser. -Separate skin thermistor. -Leighton Flexometer	Treadmill – Walking and Running Cycle Ergometer Inclinometer (tilt Table)	-3.05 -3.46 -0.61 0.20	-32.20 – 32.20 -43.70 – 43.70 -1.98 – 1.98 -2.62 – 2.62	Twenty-two physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.10 m
Johnstone et al. (2012)	-Heart Rate (beats/min). -Breathing rate (br/min).	Polar-T31. -Portable Metalyser.	Walking/Jogging	0.02 -1.19	-79.20 – 79.20 -34.40 – 34.40	Twenty (Ten for Validity) physically active males, age 21.5 ± 2.8 years, weight 71.4 ± 7.9 kg and height 1.79 ± 0.1 m

Table – 2.8 Reliability measures of Fitbit devices.

Authors	Fitbit Variables	Testing Protocol	ICC	SEM	C. I.	Mean Diff.	Sample Demographic
Diaz et al. (2015)	-Steps Taken	Walking/ Jogging	0.99	-	-	-	Twenty-three healthy adult participants, (10 males, 13 females), age 20 – 54 years, body mass index 19.60 – 29.90 kg/m^2
	-Energy Expenditure (cal.)	1.90 – 5.2 mph	0.96	-	-	-	
Takacs et al. (2013)		-Treadmill Walking					Thirty subjects (15 males. 15 females). Age 29.60 ±5.70 years and body mass index 22.70 ± 3.00 kg/m^2
	- Steps Taken	0.90 m/s	0.99	-	0.98 – 1.00	3.00	
		1.12 m/s	1.00	-	0.99 – 1.00	0.00	
		1.33 m/s	0.99	-	0.99 – 1.00	-1.00	
		1.54 m/s	0.99	-	0.99 – 1.00	-1.00	
		1.78 m/s	0.95	-	0.91 – 0.97	3.00	
	-Distance output (km)	0.90 m/s	0.99	-	0.98 – 0.99	0.00	
		1.12 m/s	0.94	-	0.90 – 0.97	-0.01	
		1.33 m/s	0.95	-	0.92 –0.98	0.00	
		1.54 m/s	0.98	-	0.97 – 0.99	-0.01	
		1.78 m/s	0.90	-	0.82 – 0.95	0.01	
Ferguson et al. (2015)	-Energy Expenditure (kcal.)	Activities of daily living for 48-hour Period	0.55	-	-	349.00	Twenty-One participants, 10 males (BMI 27.30 ± 3.20), 11 females BMI 25.50 ± 5.20, mean age 32.80 ± 10.20 years
			0.95				
	- Steps Taken			-	-	779.00	

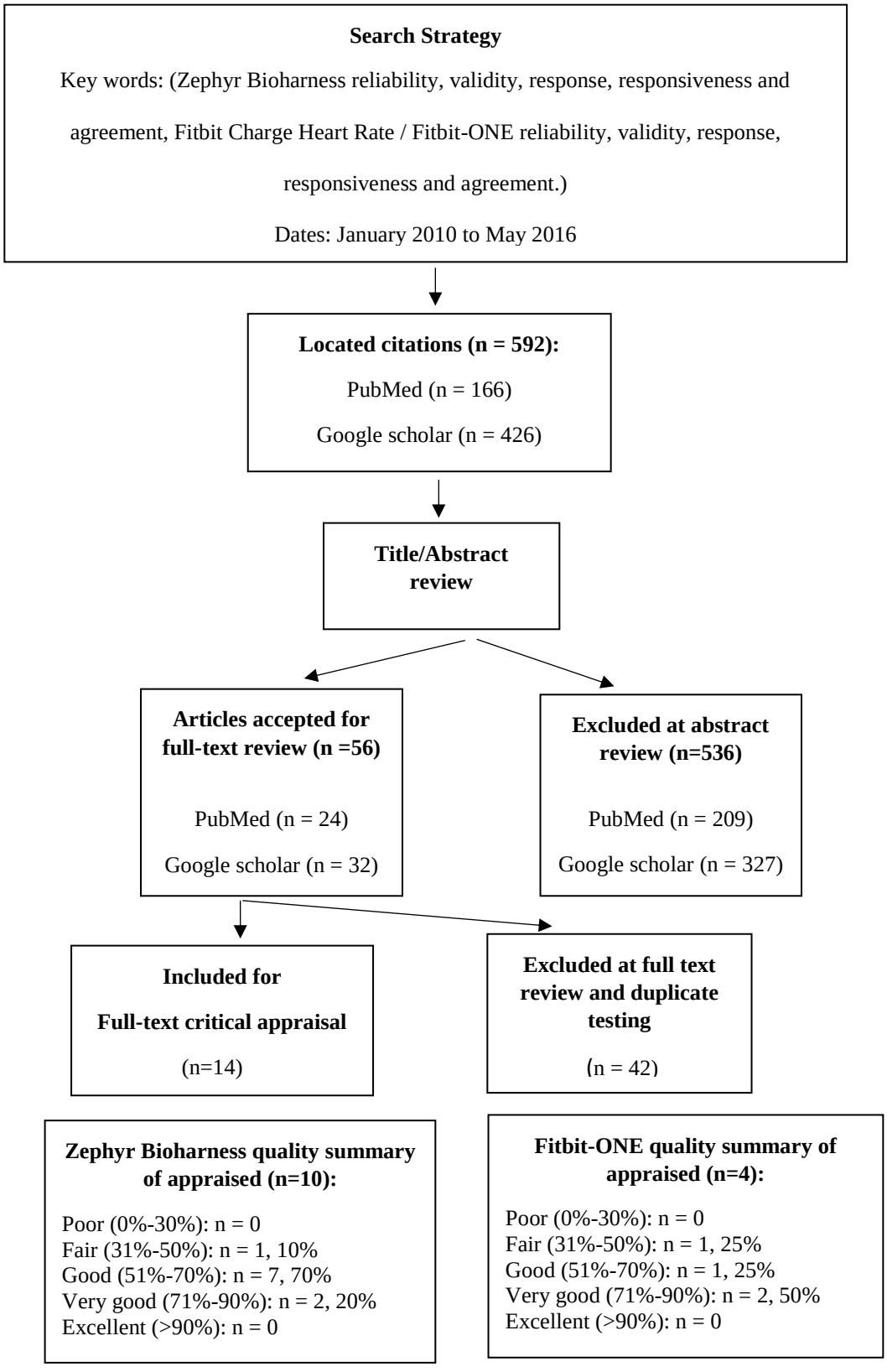
Table – 2.9 Validity Measures of Fitbit devices.

Authors	Fitbit Variables	Validity Criterion	Testing Protocol	r / rho	Mean Diff. Fitbit vs. Criterion	Sample Demographic
Diaz et al. (2015)	-Steps Taken -Energy Expenditure (cal).	-Direct Observation. -Gas Exchange Indirect Calorimeter.	Walking/Jogging 1.90 – 5.2 mph	≥ 0.97 ≥ 0.86	-3.10 – -0.30 -0.80 – 0.40	Twenty-three healthy adult participants, (10 males, 13 females), age 20 – 54 years, body mass index 19.60 – 29.90 kg/m^2
Lee et al. (2014)	-Energy Expenditure (cal.)	-Portable metabolic system	Sedentary Walking (2.5 mph) Running (5.5 mph) Moderate-vigorous activities.	0.81	-26.00	Sixty healthy participants, (Males 30, Age 28.6±6.40 years, Height 176.10 ± 5.40cm, Weight 75.40 ±9.50 kg, body mass index 24.30±2.60 kg/m^2). (Females 30, Age 24.2±4.70 years, Height 166.00 ± 7.00cm, Weight 60.30 ±8.60 kg, body mass index 21.80±2.70 kg/m^2)
Takacs et al. (2013)	-Step Count	-Observer Step Count	-Treadmill Walking	≥ 0.97	-0.40	Thirty subjects (15 males. 15 females). Age 29.60 ±5.70 years and body mass index 22.70 ± 3.00 kg/m^2
Fergusson et al. (2015)	-Energy Expenditure (kcal.) -Step Taken	BodyMedia SenseWear ActiGraph	Daily Activities for a 48-hour Period	0.76 0.99	349.00 779.00	Twenty-One participants, 10 males (BMI 27.30 ± 3.20), 11 females BMI 25.50 ± 5.20, mean age 32.80 ± 10.20 years.

Table – 2.10 Agreement Parameters of Fitbit devices.

Authors	Fitbit Variables	Criterion Measure	Testing Protocol	Agreement Bias	95% LoA	Sample Demographic
Fergusson et al. (2015)	-Energy Expenditure. (Kcal.)	BodyMedia SenseWear	Daily Activities for a 48-hour Period	-475.00	-1216.00 – 265.0	Twenty-One participants, 10 males (BMI 27.30 ± 3.20), 11 females BMI 25.50 ± 5.20, mean age 32.80 ± 10.20 years.
	-Totals Sleep time. (min.)	BodyMedia SenseWear		15.90	-51.00 – 82.90	
	-Step Count	ActiGraph		584.00	-813.00 – 1980.00	
	-Moderate to vigorous activity. (min.)	ActiGraph		65.90	-23.20 - 154.90	
Lee et al. (2014)	-Energy Expenditure (cal.)	-Portable metabolic system	Four Activities (Sedentary, walking at 2.5 mph, Running at 5.5 mph and Moderate-to-vigorous activities)	155.90	-	Sixty healthy participants, (Males 30, Age 28.6±6.40 years, Height 176.10 ± 5.40cm, Weight 75.40 ±9.50 kg, body mass index 24.30±2.60 kg/m^2). (Females 30, Age 24.2±4.70 years, Height 166.00 ± 7.00cm, Weight 60.30 ±8.60 kg, body mass index 21.80±2.70 kg/m^2)

Figure 2.1 The systematic review evidence flowchart.



CHAPTER 3:

Title of Manuscript:

Reliability of Zephyr Bioharness and Fitbit Charge Measures of Heart Rate and Activity at Rest, During the Modified Canadian Aerobic Fitness Test and Recovery.

Authors:

Goris Nazari PT, MSc (C) [a],

Joy C MacDermid PT, PhD [b-c],

Kathryn E Sinden R. Kin., PhD [d],

Julie Richardson PT, PhD [a],

Ada Tang PT, PhD [a],

[a] McMaster University, School of Rehabilitation Science, 1400 Main Street West, Hamilton, Ontario, Canada.

[b] Physical Therapy, Western University, London, Ontario, Canada.

[c] Roth McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada

[d] Department of Kinesiology and Physical Education, McGill University, 475 Pine Street, Montreal, Québec, Canada.

Corresponding Author:

Goris Nazari, McMaster University, School of Rehabilitation Science, 1400 Main Street West IAHS 403, Hamilton, Ontario, Canada. E-mail: nazarigs@mcmaster.ca

Acknowledgements:

Joy C MacDermid was supported by a Canadian Institutes of Health Research (CIHR) Chair in Gender, Work and Health and the Dr James Roth Chair in Musculoskeletal Measurement and Knowledge Translation. This work was supported by an operating grant from the Ministry of Labour - Grant Number #13-R-027. Ada Tang was supported by a personnel award from the Heart and Stroke Foundation, Ontario Provincial Office (CS I 7468).

Ethics Approval:

This study was approved by Hamilton Integrated Research Ethics Board (HiREB) of McMaster University, Hamilton, Ontario, Canada**.**

Conflict of Interest:

There are no financial gains or conflict of interest with the content of this research article.

.

Abstract

Introduction: The development of wearable physiological monitoring devices facilitates the assessment of physiologic function and movement in ways that might contribute to improved physical capacity, performance or health status. Therefore, the purpose of current study was to determine the intra-session and inter-session (test-retest) reliability of Zephyr Bioharness and Fitbit Charge variables in both healthy men and women at rest, during a sub-maximal test and throughout recovery.

Methods: Convenience and snowball sampling approaches were used to recruit sixty participants (30 females, 48 ± 15 years) and (30 males, 48 ± 15 years) from university student, staff and faculty population and MacSeniors Community Program at McMaster University. The intra-session reliability of Zephyr and Fitbit devices were assessed at rest, during the Modified Canadian Aerobic Fitness Test and throughout recovery. For inter-session reliability measures, participants were scheduled for a second session with a minimum of a 72-hour break.

Results: At rest, intra-session average heart rate (beats/min.) Intra-class correlation coefficients [ICC], Standard Error of Measurement (SEM) for Zephyr ranged from [0.94 to 0.97] (1.17 – 1.70) and [0.92 – 0.97] (1.45 – 2.10) for Fitbit Charge. During the mCAFT, the Zephyr ICCs and (SEM) ranged from 0.31 – 0.99 (1.28 – 8.10) and 0.45 – 0.99 (1.45 – 8.71) for the Fitbit Charge. Throughout the recovery, the ICCs and (SEM) ranged from 0.44 – 0.98 (1.26 – 10.47) and 0.45 – 0.98 (1.15 – 11.90) for Zephyr and Fitbit devices respectively. At rest, inter-session ICCs (SEM) for Zephyr and Fitbit ranged from 0.90 – 0.94 (1.73 – 2.37) and 0.88 – 0.94 (1.83 – 2.67) respectively. At mCAFT, the Zephyr ICCs

(SEM) ranged from 0.91 – 0.97 (3.12 – 4.64) and 0.85 – 0.98 (3.28 – 4.88) for the Fitbit. Throughout the recovery, the ICCs (SEM) ranged from 0.93 – 0.97 (2.65 – 4.66) and 0.76 – 0.91 (3.17 – 4.67) for Zephyr and Fitbit devices respectively. The number of steps taken as well as the energy expenditures (calories) recorded by the Fitbit during the mCAFT demonstrated excellent ICCs (SEM) of 0.99 (17.00) and 0.93 (9.00) respectively.

Conclusions: Both the Zephyr Bioharness and Fitbit Charge devices demonstrated excellent reliability measures at rest, during sub-maximal testing as well as throughout recovery among healthy participants across various age groups.

Introduction

Study Background:

The development of Wearable Physiological Monitoring (WPM) devices provide new ways to assess physiologic function and movement in ways that might contribute to better physical capacity, performance or overall health (Flanagan et al., 2014). The capabilities of these devices range from monitoring of physiological measures such as; heart rate and respiratory rate, to the recording of number of steps taken and energy expenditure during physical activity (Thomas, Nelson & Silverman 2005; Kim et al., 2013). Devices such as; the Zephyr BioHarness ™ (Zephyr Technology Corporation, Annapolis, MD, US), and Fitbit Charge Heart Rate (Fitbit Inc., San Francisco, CA, US), have been developed to monitor physiological and physical activity measures (Zephyr Technology Manual 2012; Fitbit Charge HR Manual 2015).

Heart rate monitoring has various advantages. It has become one of the most frequently used methods of evaluating exercise intensity (Achten & Jeukendrup 2003). In addition, heart rate monitoring has also been proposed as a method to determine anaerobic threshold (Conconi et al., 1982). Over the last several decades, data obtained from the linear relationship between heart rate and oxygen consumption (VO_2) has been used to estimate maximal oxygen uptake VO_{2max} (Astrand & Ryhming 1954). Manual pulse palpation provides inaccurate results and the use of electrocardiogram or Holter monitoring to assess heart rate requires multiple wires connecting electrodes with data processing units and these technical requirements limit how and where the devices can be used (Jovanov et al., 2002; Bianchi et al., 2013). Both the Zephyr Bioharness (ZB) and the Fitbit Charge Heart Rate

(FC-HR) devices provide wireless ambulatory physiological measures of heart rate. Furthermore, several studies have taken the cheaper and an easier alternative to investigate Energy Expenditure estimates from heart rate measures (Luke et al., 1997; Treuth, Adolph & Butte 1998).

Physical inactivity imposes a major economic burden on the Canadian health care system. It has been reported that in 2009, physical inactivity accounted for $ 6.70 billion as direct health care costs in Canada: hospital, drug and physician care expenditures, as well as indirect expenditures; economic output lost due to illness, injure-related work disability or premature death (Janssen 2012). Physical inactivity is an independent and modifiable risk factor for chronic disability and it has been demonstrated that regular physical activity has been associated with health improvements in many populations (Alwan 2010). In an attempt to promote regular physical activity, WPM devices recording the number of steps taken and levels energy expenditure have been shown to be effective in increasing physical activity levels (Heath et al., 2012).

An essential requirement of any device that is used to assess an individual's status is that it should be reliable. Reliability is referred to as the overall consistency of a device (De Vet et al., 2006). It deals with the extent to which repeated measurements in stable participants yield similar results (Carlson et al., 2009). Recent studies have only examined the test-retest reliability of ZB. Johnstone et al. (2012) and Johnstone et al. II (2012), have both reported excellent test-retest reliability measures (ICC $\geq$ 0.90), in ten healthy young male participants (21.5 $\pm$ 2.8 years) and (20.5 $\pm$ 2.1 years) during a treadmill testing protocol. The third study conducted by Rawstorn et al. (2015), reported the reliability

measures in two phases. The first phase tested 10 people who were recreationally active including six males and four females, (26.68 ± 3.26 years) during a treadmill running test, and then, in the second phase test five males and three females with atrial fibrillation, (69.68 ±9.53 years) during a treadmill test, a cycle ergometer and during stimulated activities of daily living (sweeping and vacuuming). Both phases demonstrated that ZB was able to identify excellent reliability measures (ICC > 0.90). Limitations in the studies to date for the ZB include the focus on only young males, small sample sizes and measure of reliability only during activity. Furthermore, these findings do not provide any information regarding the stability (intra-session reliability) of the ZB device within a single session. As for FC-HR device, there have been no studies that have reported its reliability measures.

Therefore, studies with larger sample sizes, investigating both the intra-session and inter-session reliability of ZB and FC-HR in both healthy men and women at rest, during a sub-maximal test and throughout recovery, is warranted.

Study Research Question:

The overall aim of this study was to determine the reliability of ZB and FC-HR devices. Specifically: 1) To determine the intra-session and inter-session reliability of ZB and FC-HR heart rate variables at rest, during modified Canadian Aerobic Fitness Test (a submaximal test) and throughout recovery. 2) To establish the inter-session reliability of FC-HR variables of steps taken and energy expenditure, during the submaximal test. 3) To determine which criterion measure of heart rates (the first measurement, the first three

measurements, the first-five measurements or all measurements), would provide a more consistent inter-session reliability measure of the heart rate by the FC-HR device.

Methods

Sampling and Recruitment:

After securing the ethical approval for this study through the Hamilton Integrated Research Ethics Board (No. 0825), a total of sixty participants from three age groups were recorded: 20-39 (n=10 males and n=10 females), 40-59 (n=10 males and n=10 females) and 60-69 (n=10 males and n=10 females) years. Stratified convenience and snowballing sampling approaches were used from university student, staff and faculty population and MacSeniors Community Program at the McMaster Physical Activity Centre of Excellence.

Inclusion Criteria:

Healthy participants both males and females 20 – 69 years of age, able to read, write and communicate in English were eligible to take part in this study (Canadian Society for Exercise Physiology 1998). Participants completed a self-reported Physical Activity Readiness Questionnaire (PAR-Q) and were eligible to take part if had answered "No" to all seven PAR-Q questions.

Exclusion Criteria:

Participants were excluded from study if they had answered "Yes" to any of seven PAR-Q questions (Canadian Society for Exercise Physiology 1998) (Appendix C).

Sample Size Calculation for Relative Reliability Hypothesis Testing:

The sample size was based on the null hypothesis value of test-retest reliability ICC = 0.80 and the expectation of obtaining an (ICC) of 0.90 in this study (Donner & Eliasziw 1987; Johnstone et al., 2012; Johnstone et al., 2012b). Based on the calculations, a sample size of 60 participants was required (Appendix A).

Wearable Physiological Monitoring devices:

Zephyr BioHarness

The Zephyr BioHarness (ZB) is a U.S. FDA-approved wireless ambulatory physiological monitoring device capable of capturing and the transmission of wide range of physiological data on the wearer including; heart rate, respiratory rate, estimated core temperature, posture and activity levels (Zephyr Technology Manual 2012). The device includes a BioModule that snaps onto an adjustable chest strap which is suitable for both men and women and is worn at the lower sternum. The device measures heart rate by recording cardiac electrical impulses using conductive fabric skin electrodes and reports beats per minute (beats/minute) (Zephyr Technology Manual 2012). The cardiac impulses are then sent to and processed by the BioModule. The BioModule can capture data for 26 hours when charged for three hours and can log up to 20 days of data.

Fitbit Charge Heart Rate

The Fitbit Charge Heart Rate (FC-HR) is a flexible, elastomer wristband with a surgical-grade stainless steel buckle which would fit wrists 13.7 to 22.1 cm in circumference. It

captures physiological measures; wrist-based heart rate, number of steps taken and caloric (energy) expenditure (Fitbit Charge HR Manual 2015). The device is sweat, rain and splash proof and is powered by a Lithium-polymer battery and can store data up to 30 days and last for 5-days upon a two-hour charge.

Modified Canadian Aerobic Fitness Test:

The Modified Canadian Aerobic Fitness Test (mCAFT) is an eight-stage step-test. Based on age and gender, individuals' initial stepping stage and 85% of maximum heart rate (ceiling post-exercise heart rate) were determined (Canadian Society for Exercise Physiology 1998). Throughout the initial stepping stage and based on a predetermined cadence (foot-plants/minutes), individuals completed their first three-minute stepping session on a double 20.3cm step stool. Heart rates were measured at the end of stepping session. The measured heart rates were compared with the ceiling post-exercise heart rates and a heart rate value less than ceiling post-exercise heart rate, indicated the need to carry out a second three-minute stepping session (Canadian Society for Exercise Physiology 1998). Multiple three-minute stepping sessions were performed in order to achieve a heart rate value that equalled or exceeded the ceiling post-exercise heart rate. Once this was achieved, the test was completed. The mCAFT was chosen because it is standardized, feasible and replicates a functional task; stair climbing (Carlson et al., 2009). In addition, it is a well-established reliable and valid sub-maximal test and provides an estimate of $VO_{2\text{-}max}$ (Weller et al. 1994; Weller & Corey 1998).

Study Design:

The mCAFT Preliminary Participant Instructions Form (Do's and Don'ts prior to first session – Appendix E) was emailed to all the potential participants 48 hours prior to their first visit. During the first session, the study was explained in detail to participants and they were asked to complete the PAR-Q and sign a Consent Form. Then the screening procedures; resting heart rate, resting blood pressure, height (m) and body weight (kg) measures, were carried out (Canadian Society for Exercise Physiology 1998). Next, was the fitting of the ZB and the FC-HR devices. Participants were then asked to be seated for a period of ten minutes –that served as rest period. Following this ten-minute rest period, the mCAFT was administered. The mCAFT termination was followed by a ten-minute recovery period where the participants were asked to be seated again.

For the inter-session reliability measures, the participants were scheduled for a second visit following a minimum of 72 hours break and the exact same procedures were carried out (Johnstone et al., 2012). The minimum 72-hours break was chosen based on the previous reliability studies.

Measurement Protocol:

To determine the intra-session heart rate reliability of these two devices, ZB and FC-HR were used to record participants' heart rate at rest, during the mCAFT and throughout the recovery during the first session. At Rest, heart rate measurements were taken at 30-second intervals over the 10-minute period for a total 20 measurements. During the mCAFT a total of 24 measures for a duration of twelve minutes at 30-second consecutive intervals, and

throughout the recovery a total of 20 measures for a period of ten minutes at 30-second consecutive intervals were recorded. To determine the inter-session reliability of ZB and FC-HR heart rate variable, the same 30-second heart rate interval recordings were carried out at rest, during the mCAFT as well as at recovery during the first and second sessions minimum of a 72-hour break between sessions. In addition, the FC-HR device was also used to record the total number of steps taken as well as energy expenditure during the mCAFT for both the sessions.

Statistical Analysis

Demographic characteristics of the sample stratified by gender; women and men, including age, height, weight, body mass index were described using means, standard deviations, minimum and maximum scores.

The heart rate was monitored and recorded by ZB and FC-HR devices throughout the rest, during the mCAFT and recovery. The ZB data file was exported into Microsoft Excel 2016, and was stratified into 30-second time intervals. The FC-HR data was collected through the Fitbit Dashboard software using The Fitbit App, at 30-second time intervals. FC-HR steps taken and energy expenditure data files were extracted using the recommended software put forward by the Fitbit manufacturers similar to how a consumer would use the software. For the ZB and FC-HR intra-session heart rate reliability, means (M), mean differences (Mean Differences), Standard Error of Measurement (SEM) and the Intra-Correlation Co-efficient (ICC) including the 95% confidence intervals between all the 30-second consecutive intervals at rest, during the mCAFT and throughout the recovery, during the

first session were assessed. For the ZB and FC-HR inter-session heart rate reliability (M), (Mean Differences), (SEM) and the (ICC) including the 95% confidence intervals between all the 30-second consecutive intervals at Rest, mCAFT and Recovery at the first and the second sessions were assessed.

For the FC-HR inter-session number of steps taken and energy expenditure reliability measures, (M), (Mean Differences), (SEM) and (ICC) including the 95% confidence intervals only during the mCAFT were examined. In addition, the (M), (Mean Differences), (SEM) and the (ICC) including the 95% confidence intervals of the first, first three, first five and all the measures of heart rate recorded by FC-HR at rest, mCAFT and recovery were reported to assess which set of measures by FC-HR provide more consistent inter-session heart rate reliability. For all the reliability measures, the single measures of ICC with a two-way mixed effect model and an absolute agreement type were reported. An ICC < 0.40 indicated poor, $0.40 \leq$ ICC < 0.75 indicated fair to good and ICC ≥ 0.75 indicated excellent reliability (Rosner 2005). Analyses were performed using IBM SPSS Statistics software version 22.0.

Results

Sample

Sixty (30 females, 30 males) healthy participants took part in this study. Table – 3.1 displays the demographic characteristics; age, height, weight, body mass index, stratified by gender; females and males.

Intra-session Reliability of ZB and FC-HR for heart rate Variable:

At rest, intra-session heart rate ICCs (SEM) for ZB ranged from 0.94 to 0.97 (1.17 – 1.70) and 0.92 – 0.97 (1.45 – 2.10) for FC-HR (Table – 3.2). During the mCAFT, the ZB ICCs (SEM) ranged from 0.31 – 0.99 (1.28 – 8.10) and 0.45 – 0.99 (1.45 – 8.71) for the FC-HR (Table – 3.3). As anticipated, the lowest intra-session ICCs and the largest mean differences registered were during the first minute of the mCAFT. The ICCs (Means) of 0.31 and (71.55 – 102.90) and 0.75 (102.90–111.60) were recorded by ZB for the 1st and the 2nd thirty-second intervals respectively. As for the FC-HR, the 1st and the 2nd measures recorded ICCs (Means) of 0.45 (72.00 – 89.40) and 0.55 (89.40 – 104.00) respectively. Throughout the recovery, the ICCs (SEM) ranged from 0.44 – 0.98 (1.26 – 10.47) and 0.45 – 0.98 (1.15 – 11.90) for ZB and FC-HR devices respectively (Table – 3.4). The lowest intra-session ICCs and the largest mean differences were observed during the first three measures (1.5 minutes into the recovery phase). The ICCs (Means) of 0.44 (151.50–132.47), 0.44 (132.47–113.00) and 0.71 (113.02-103.20) were recorded by ZB for the 1st, 2nd and 3rd measures respectively. As for the FC-HR, the 1st, 2nd and 3rd measures registered ICCs (Means) of 0.50 (150.88 – 131.27), 0.55 (131.27 – 113.78) and 0.70 (113.78 – 103.82) respectively.

Inter-session Reliability of ZB and FC-HR for heart rate Variable:

At rest, inter-session ICCs (SEM) for ZB and FC-HR ranged from 0.90 – 0.94 (1.73 – 2.37) and 0.88 – 0.94 (1.83 – 2.67) respectively (Table – 3.5). At mCAFT, the ZB ICCs (SEM) ranged from 0.91 – 0.97 (3.12 – 4.64) and 0.85 – 0.98 (3.28 – 4.88) for the FC-HR (Table

– 3.6). Throughout the recovery, the ICCs (SEM) ranged from 0.93 – 0.97 (2.65 – 4.66) and 0.76 – 0.91 (3.17 – 4.67) for ZB and FC-HR devices respectively (Table – 3.7).

Inter-session Reliability of Number of Steps Taken and Energy Expenditure:

The number of steps taken as well as the energy expenditures recorded by the FC-HR during the mCAFT only, demonstrated excellent ICCs (SEM) of 0.99 (17.00) and 0.93 (9.00) respectively (Table – 3.8). The number of steps taken and the energy expenditures were not observed at rest or throughout recovery.

Combined Inter-session reliability of FC-HR heart rate variable:

At rest, the first measure and all the 20 measures provided very similar ICCs; 0.94, 0.95, and SEM; 1.98, 2.09, respectively (Table – 3.9). During the mCAFT, the First measure yielded a higher ICC score; 0.91, as compared to the ICC of all the twenty measures; 0.81, however similar SEM were reported (Table – 3.9). Throughout the recovery ICCs and (SEM) of 0.80 (4.67) and 0.90 (4.10) were reported for the first and all twenty measures respectively (Table – 3.9).

Discussion

This study established that high reliability can be expected in healthy participants of different age groups, both within and across test sessions, when using either the ZB and FC-HR to assess heart rate variables at rest, during the mCAFT and throughout recovery. As anticipated, stable responses cannot be expected during times of elements of a test

protocol where changes in activity are happening. Furthermore, estimates of activity including the number of steps taken and energy expenditure also demonstrated high reliability.

In our previous systematic review of the measurement properties of Zephyr and the Fitbit devices, we indicated the importance of considering the device, the population, activity and context, when interpreting measurement properties. The context of our study and how study measurements were taken should be considered when interpreting our reliability measures since we used a submaximal fitness test which introduced a metabolic stress to the system and was expected to induce physiological changes; but reliability should be determined under stable conditions. We did this because it was important to learn that the devices provided reliable results not just when people are at rest, but more importantly during activity as this is their main purpose. Thus we have divided our testing into different intervals. We might anticipate that after a period of appropriate acclimatization that rest would represent a stable condition, that the submaximal fitness test would introduce substantial perturbation that would be reflected in physiological changes and low reliability coefficients, and that recovery would be characterized by resumption to a stable condition. We anticipate that across the testing, the most unstable intervals would be those where there is transitioning between physical activity and rest or vice versa. Measuring the consistency of a device under unstable conditions would yield lower reliability values. When considering this, our high reliability coefficients in more stable portions of the test; and lower reliability coefficients during times of transition were both anticipated and reflect the context of our testing.

Since the mCAFT was administered after a ten-minute period of rest, where participants had already achieved a stable heart rate, the lower ICCs, larger SEM and mean difference were expected at the beginning of mCAFT. The mCAFT is a multi-stage progressively demanding stepping test where a predetermined cadence (foot-plants/minute) is assigned to each stage. As the testing progressed, the stepping cadence increased as well. However, this increase was gradual. The stepping cadence ranges from 66 – 144 steps/minute for stages 1 – 6 for men and 1 – 7 for women. This steady steps/minute increase throughout 6 or 7 stages contributed to higher ICCs and lower heart rate mean differences during the mCAFT throughout the 3rd to 24th measures. The recovery phase was initiated when participants achieved their ceiling post-exercise heart rate at the end of a given three-minute stepping stage. The maximum post-exercise heart rate was referred to the 85% of participants' heart rate – maximum. (i.e. 85% of 220 – Age). The lower ICCs, larger SEM and mean differences were again expected because participants were working at 85% of their HR-max for a certain given period of time and then immediately asked to sit (recovery phase) when the stepping stage was over. Therefore, we did not expect their heart rates to be stable at the very beginning of the Recovery phase. However, throughout the 4th to 20th measure (1.5 – 10.0 minutes) higher ICCs, smaller SEM and lower mean differences were recorded by both the devices.

Others have also reported excellent intersession reliability for the ZB at rest; 2-minutes of upright standing before exercise (Johnstone et al., 2012b). This study was conducted on a sample of ten physically active males, age 20.5 ± 2.1 years. The narrow age range and similar physical activity levels of this group might be expected to contribute to a more

uniform physiological response. Our study adds to this literature as we included a more diverse age range and both genders. Similarly, ICCs exceeding 0.90 with SEM 4.77, have been reported for the ZB during Treadmill testing (Rawstorn et al., 2015). However, the inclusion of cardiac patients makes comparison with our ZB findings difficult. In addition, since our study was the first to determine the reliability of FC-HR at rest, mCAFT and recovery phases, no directly comparable studies were found.

It is common for people to use wearable sensors to measure their overall activity in terms of the total number of steps taken and energy expenditure. We found excellent intersession reliability for these measures using the FC-HR. This concurs a prior study that found excellent reliability with an earlier version of Fitbit – Fitbit-Flex, when evaluated in healthy young adults during treadmill testing (Diaz et al., 2015). Since a different version of the device was used, our findings can not be directly compared, however it does provide consistent findings regarding the reliability parameters of Fitbit.

One of the issues we wished to examine in this study was whether it made a difference how many repeated measures were used to establish the criterion measure. This is important since it could provide evidence for users about how they should be using their device to provide more reliable indicators of their physiological status. For example, taking multiple measurement is common when assessing grip strength or blood pressure. Therefore, we examined the intersession reliability of the FC-HR device based on the first measure, the first three, first-five and all the measures of heart rate during the rest, mCAFT and recovery phases. The criterion measure was established at a single measure, since the first measure at rest, mCAFT and recovery provided excellent reliability.

Strengths of this study included the fact that we used multiple devices during standardized submaximal fitness test; that we included a range of ages and both genders and that we sampled a greater number of participants that have been previously reported on. Despite this, limitations inherent in our approach should be considered when interpreting our findings. We studied the reliability of these devices during human performance and did not evaluate the performance of the sensor to provide accurate outputs against a gold standard (calibration). Neither the validity nor the levels of agreement between the ZB and FC-HR device were assessed in this study. A device could be reliable but not measure what it is intended to measure, i.e. be invalid (Brazier & Deverill 1999). In addition, a valid instrument may still provide a lack of agreement despite high correlations indicating the presence of bias (Bunce 2009). This occurs when one instrument consistently reads lower or higher than the other (Bunce 2009). Therefore, assessing both the validity and levels of agreement between the ZB and FC-HR devices are warranted. Finally, since it was impossible to define exactly how long to rest or activity to reach a steady-state we did not specify ahead of time where we expected will high reliability coefficients that reflect reliability, versus low correlations that would be expected when the device was actually measuring true change.

Conclusion

The reliability of two Wearable Monitoring Devices (Zepher Bioharness and Fitbit Charge) provided excellent reliability measures of heart rate and activity (steps taken or energy expenditure) during a standardized submaximal fitness test in a sample of sixty healthy

male and female participants of various age groups throughout three phases; rest, mCAFT testing and recovery.

References:

Achten, J., and Jeukendrup, A., (2003). Heart Rate Monitoring: Applications and Limitations. *Sport Medicine,* 33(7), 517-538.

Alwan, A. Global status report on non-communicable diseases (2010). 1st ed. Italy:*World Health Organization*; 2011.

Astrand, P. O., and Ryhming, I,. (1954). Nomogram for calculations of aerobic capacity from pulse rate during submaximal work [abstract]. *J Appl Physiol,* 7, 218.

Bianchi, W., Dugas, A. F., Hsieh, Y. H., Saheed, M., Hill, P., Lindauer C., …Rothman R. E. (2013). Revitalizing a vital sign: Improving detection of tachypnea at primary triage. *Annals of Emergency Medicine,* 61 (1), 37 – 43.

Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related qualityof life: learning from psychometrics. *Health Economics,* 8, 41–51.

Bunce, C., (2009). Correlation, Agreement, and Bland–Altman Analysis: Statistical Analysis of Method Comparison Studies. *Am J Ophthalmol,*148(1), 4-6.

Canadian Society for Exercise Physiology (1998). *The Canadian Physical Activity, Fitness and Lifestyle Appraisal* (2nd ed.). Ottawa, ON.

Carlson, N. R., Buskist, W., Heth, C. D. and Schmaltz, R. (2009*). Psychology : The science of behaviour* (4th Canadian ed. ed.). Toronto: Pearson.

Conconi, F., Ferrari, M., Ziglio, PG., Droghetti, P. and Codeca, L. (1982). Determination of the anaerobic threshold by a non-invasive field test in runners. *J Appl Physiol*, 52 (4): 869-73.

Diaz, K.M., Krupka, D.J., Chang, M.J., Peacock, J., Ma, Y., Goldsmith, J., … Davidson, K.W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, 185, 138–140.

Donner, A., and Eliasziw, M. (1987). Sample Size Requirements for Reliability Studies. *Statistics in Medicine*, 6, 441-448.

Fitbit charge HR Wireless Heart Rate + Activity Wrist Band (2015, January*). Fitbit charge HR$_{TM}$ Product Manual Version 1.0.* Retrieved from https://www.fitbit.com/ca/chargehr

Flanagan, S.D., Comstcok, B.A., Dupont, W.H., Sterczala, A.R., Looney, D.P., Dombrowski, D.H., Mcdermott, D.M., Byrce, A., …, Kraemer, W.J. (2014). Concurrent Validity of the Armour-39 heart Rate Monitor Strap. *Journal of Strength and Conditioning Research*, 28(3) 870 – 873.

Heath, G. W., Parra, D. C., Sarmiento, O. L., Andersen, L. B., Owen, N., Goenka, S., Montes, F. and Brownson, R.C. (2012). Lancet Physical Activity Series Working Group: Evidence-based intervention in physical activity: lessons from around the world. *Lancet*, 380 (9838),272–281.

De Vet, H.C., Terwee, C. B., Knol. D.L. & Bouter, L.M., (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology* 59, 1033–1039.

Janssen, I. (2012). Health care costs of physical inactivity in Canadian adults. *Appl. Physiol. Nutr. Metab.*, 37: 803 – 806.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T., Mitchell A.C. and Garrett, A.T. (2012). Field based reliability and validity of the Bioharness^{TM} multivariable monitoring device. *Journal of Sports Science and Medicine* 11, 643-652.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012b). Bioharness™ multivariable monitoring device: Part II: Reliability. *Journal of Sports Science and Medicine* 11(3), 409- 417.

Jovanov, E., O'Donnel, A., Morgan, A., Priddy, B. & Hormigo, R. (2002). Prolonged telemetric monitoring of the heart rate variability using wireless intelligent sensors and mobile getaway. Proceedings of the Second Joint EMBS/BMES Conference, Houston, TX, USA.

Kim, J.H., Roberge, R., Powell, J.B., Shafer, A.B., and Williams, J.W. (2013). Measurement Accuracy of Heart Rate and Respiratory Rate during Graded Exercise and Sustained Exercise in the Heat Using the Zephyr BioHarness™. *Int J Sports Med,* 34, 497–501.

Luke, A., Maki, K.C., Barkey, N., Cooper R. and McGee D. (1997). Simultaneous monitoring of heart rate and motion to assess energy expenditure. *Med Sci Sports Exerc*, 29 (1): 144-148.

Rawstorn, J.C., Gant, N., Warren, I., Doughty, R.N., Lever, N., Poppe, K.K. and Maddison, R. (2015). Measurement and data transmission: Validity of a multi-biosensor

system for real-time remote exercise monitoring among cardiac patients. *JMIR Rehabil Assist Technol,* 2 (1): e2.

Rosner, B. (2005). Fundamentals of biostatistics (6th ed.). Boston: Duxbury Press.

Thomas, J.R., Nelson, J.K., Silverman, S.J., (2005). Research Methods in Physical Activity. 5th Edition, Champaign IL: Human Kinetics Publishers.

Treuth, M.S., Adolph, A.L., and Butte, N.F. (1998). Energy expenditure in children predicted from heart rate and activity against respiration calorimetry. *Am J Physiol,* 275: E12-8.

Weller, I.M., Thomas, S.G., Gledhill, N., Paterson, D. and Quinney, A. (1994). A study to validate the modified Canadian Aerobic Fitness Test. *Can J Appl Physiol*, 20, 211-221.

Weller, I.M. and Corey, P.N. (1998). A study of the reliability of the Canada Fitness Survey questionnaire. *Med Sci Sports Exerc,* 30, 1530-1536.

Zephyr Technology (2012, September 12). *BioHarness 3.0 User ManualBioHarness 3.0 User Manual*. Retrieved from http://zephyranywhere.com/products/bioharness-3/

Table – 3.1 Demographics

	Women (n = 30)				**Men (n = 30)**			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Age (yrs.)	48.00	15.00	23.00	68.00	48.00	15.00	21.00	68.00
Height (m)	1.70	0.05	1.60	1.77	1.78	0.06	1.68	1.88
Weight (kg)	69.00	11.00	50.00	100.00	79.00	8.50	64.00	97.00
Body Mass Index (kg/m^2)	24.00	3.50	18.00	34.00	25.00	2.30	20.00	31.00
HR-max Age-related	146.50	13.00	129.00	167.00	146.10	12.85	129.00	169.00

Table – 3.2 Intra-session reliability of Zephyr Bioharness and Fitbit Charge heart rate variables at Rest.

Measure	Time (min: sec)	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	71.19– 71.32	-0.13	1.28	0.97	0.95 – 0.98
2nd	0:30 – 1.00	71.32– 71.62	-0.3	1.30	0.97	0.95 – 0.98
3rd	1:00 – 1:30	71.62– 71.67	-0.05	1.26	0.97	0.95 – 0.98
4th	1:30 – 2:00	71.67– 71.72	-0.05	1.60	0.95	0.93 – 0.97
5th	2:00 – 2:30	71.72– 71.30	0.42	1.50	0.96	0.93 – 0.97
6th	2:30 – 3:00	71.30– 71.57	-0.27	1.41	0.96	0.94 – 0.97
7th	3:00 – 3:30	71.57– 71.62	-0.05	1.70	0.94	0.91 – 0.96
8th	3:30 – 4:00	71.62– 71.50	0.12	1.65	0.95	0.92 – 0.97
9th	4:00 – 4:30	71.50– 71.73	-0.23	1.40	0.96	0.94 – 0.98
10th	4:30 – 5:00	71.73– 71.35	0.38	1.54	0.95	0.93 – 0.97
11th	5:00 – 5:30	71.35– 71.63	-0.28	1.63	0.95	0.92 – 0.97
12th	5:30 – 6:00	71.63– 71.67	-0.04	1.56	0.95	0.92 – 0.97
13th	6:00 – 6:30	71.67– 71.32	0.35	1.68	0.94	0.91 – 0.96
14th	6:30 – 7:00	71.32– 71.48	-0.16	1.48	0.96	0.93 – 0.97
15th	7:00 – 7:30	71.4 – 71.35	0.13	1.56	0.95	0.92 – 0.97
16th	7:30 – 8:00	71.35– 71.58	-0.23	1.47	0.95	0.92 – 0.97
17th	8:00 – 8:30	71.58– 71.85	-0.27	1.30	0.96	0.94 – 0.98
18th	8:30 – 9:00	71.85– 71.80	0.05	1.34	0.96	0.94 – 0.98
19th	9:00 – 9:30	71.80– 71.55	0.25	1.17	0.97	0.96 – 0.98
20th	9:30 – 10:00	71.55– 71.55	0.00	1.22	0.97	0.95 – 0.98

Table – 3.2 Intra-session reliability of Zephyr Bioharness and Fitbit heart rate variables at Rest.

Measure	Time (min: sec)	FitBit – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	71.70 – 71.90	-0.20	1.55	0.97	0.94 – 0.99
2nd	0:30 – 1.00	71.90 – 72.00	-0.10	1.60	0.96	0.93 - 0.97
3rd	1:00 – 1:30	72.00 – 71.70	0.30	1.73	0.95	0.91 – 0.97
4th	1:30 – 2:00	71.70 – 71.90	-0.20	1.88	0.94	0.89 – 0.96
5th	2:00 – 2:30	71.90 – 71.33	0.57	1.74	0.95	0.91 – 0.97
6th	2:30 – 3:00	71.33 – 71.53	-0.20	1.78	0.94	0.90 – 0.96
7th	3:00 – 3:30	71.53 – 71.78	-0.25	2.10	0.92	0.87 – 0.95
8th	3:30 – 4:00	71.78 – 71.95	-0.17	1.90	0.94	0.90 – 0.96
9th	4:00 – 4:30	71.95 – 71.87	0.08	1.94	0.93	0.89 – 0.96
10th	4:30 – 5:00	71.87 – 71.67	0.20	1.75	0.94	0.91 – 0.97
11th	5:00 – 5:30	71.67 – 71.87	-0.20	2.05	0.92	0.87 – 0.95
12th	5:30 – 6:00	71.87 – 71.65	0.22	1.97	0.93	0.88 – 0.96
13th	6:00 – 6:30	71.65 – 71.82	-0.17	1.93	0.93	0.89 – 0.96
14th	6:30 – 7:00	71.82 – 71.63	0.19	2.10	0.93	0.88 – 0.96
15th	7:00 – 7:30	71.63 – 71.63	0.00	2.00	0.95	0.92 – 0.97
16th	7:30 – 8:00	71.63 – 71.85	-0.22	1.90	0.93	0.89 – 0.96
17th	8:00 – 8:30	71.85 – 71.92	-0.07	1.45	0.96	0.93 – 0.98
18th	8:30 – 9:00	71.92 – 72.02	-0.10	1.67	0.95	0.92 – 0.97
19th	9:00 – 9:30	72.02 – 72.00	0.02	1.56	0.96	0.93 – 0.97
20th	9:30 – 10:00	72.00 – 71.55	0.45	1.83	0.94	0.90 – 0.96

Table – 3.3 Intra-session reliability of Zephyr Bioharness *and Fitbit Charge* heart rate variables during mCAFT.

Measure	Time (min: sec)	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	71.55 – 102.90	-31.35	8.10	0.31	0.2 – 0.92
2nd	0:30 – 1.00	102.90–111.60	-8.70	6.63	0.75	-0.04 – 0.92
3rd	1:00 – 1:30	111.60–118.40	-6.80	5.34	0.87	0.05 – 0.96
4th	1:30 – 2:00	118.40–120.80	-2.40	3.27	0.99	0.89 – 0.98
5th	2:00 – 2:30	120.80–122.90	-2.10	2.55	0.98	0.93 – 0.99
6th	2:30 – 3:00	122.90–124.60	-1.70	2.54	0.98	0.96 – 0.99
7th	3:00 – 3:30	118.10–114.80	3.30	3.87	0.93	0.80 – 0.97
8th	3:30 – 4:00	114.80–124.30	-9.50	6.64	0.81	-0.04 – 0.95
9th	4:00 – 4:30	124.30–128.40	-4.10	4.11	0.93	0.63 – 0.97
10th	4:30 – 5:00	128.40–130.30	-1.90	2.80	0.97	0.93 – 0.98
11th	5:00 – 5:30	130.30–132.70	-2.40	2.10	0.98	0.71 – 0.99
12th	5:30 – 6:00	132.70–134.30	-1.60	1.75	0.99	0.95 – 0.99
13th	6:00 – 6:30	127.70–124.50	3.20	3.71	0.92	0.70 – 0.96
14th	6:30 – 7:00	124.50–131.70	-7.20	5.10	0.82	-0.04 – 0.95
15th	7:00 – 7:30	131.70–138.00	-6.30	4.53	0.86	-0.03 – 0.96
16th	7:30 – 8:00	138.00–141.80	-3.80	3.23	0.92	0.26 – 0.97
17th	8:00 – 8:30	141.80–143.90	-2.10	1.96	0.97	0.77 – 0.99
18th	8:30 – 9:00	143.90–146.60	-2.70	2.83	0.95	0.74 – 0.98
19th	9:00 – 9:30	137.30–128.60	8.70	5.81	0.83	-0.02 – 0.97
20th	9:30 – 10:00	128.60–139.00	-10.40	6.90	0.79	-0.05 – 0.96
21th	10:00 – 10:30	139.00–145.30	-6.30	4.44	0.92	-0.01 – 0.98
22nd	10:30 – 11:00	145.30–148.20	-2.90	2.69	0.97	0.65 – 0.99
23rd	11:00 – 11:30	148.20–149.8	-1.60	1.28	0.99	0.47 – 0.99
24th	11:30 – 12:00	149.80–151.50	-1.70	2.02	0.98	0.90 – 0.99

Table – 3.3 Intra-session reliability of Zephyr Bioharness and Fitbit Charge heart rate variables during mCAFT.

Measure	Time (min: sec)	FitBit – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	72.00 – 89.40	-17.40	7.43	0.45	0.02 – 0.87
2nd	0:30 – 1.00	89.40 – 104.00	-14.60	8.71	0.55	-0.07 – 0.85
3rd	1:00 – 1:30	104.00 – 114.63	-10.63	7.57	0.75	-0.07 – 0.93
4th	1:30 – 2:00	114.63 – 119.45	-4.82	5.53	0.90	0.67 – 0.96
5th	2:00 – 2:30	119.45 – 121.20	-1.75	3.83	0.96	0.93 – 0.98
6th	2:30 – 3:00	121.20 – 122.63	-1.43	2.52	0.98	0.97 – 0.99
7th	3:00 – 3:30	116.00 – 110.14	5.86	5.14	0.89	0.28 – 0.97
8th	3:30 – 4:00	110.14 – 118.76	-8.62	6.20	0.83	-0.04 – 0.96
9th	4:00 – 4:30	118.76 – 124.22	-5.46	4.20	0.94	0.07 – 0.99
10th	4:30 – 5:00	124.22 – 127.60	-3.38	3.52	0.96	0.80 – 0.99
11th	5:00 – 5:30	127.60 – 129.84	-2.24	2.50	0.98	0.92 – 0.99
12th	5:30 – 6:00	129.84 – 132.67	-2.83	2.64	0.98	0.78 – 0.99
13th	6:00 – 6:30	125.00 – 120.87	4.13	4.85	0.87	0.35 – 0.96
14th	6:30 – 7:00	120.87 – 127.39	-6.52	4.82	0.86	-0.03 – 0.96
15th	7:00 – 7:30	127.39 – 134.05	-6.66	4.78	0.86	-0.03 – 0.97
16th	7:30 – 8:00	134.05 – 138.56	-4.51	3.38	0.93	0.01 – 0.98
17th	8:00 – 8:30	138.56 – 141.51	-2.95	2.50	0.96	0.42 – 0.99
18th	8:30 – 9:00	141.51 – 144.00	-2.49	2.28	0.97	0.64 – 0.99
19th	9:00 – 9:30	135.50 – 124.88	10.62	6.80	0.77	-0.03 – 0.96
20th	9:30 – 10:00	124.88 – 135.78	-10.90	7.00	**0.75**	-0.06 – 0.96
21th	10:00 – 10:30	135.78 – 143.00	-7.22	5.47	0.88	-0.03 – 0.98
22nd	10:30 – 11:00	143.00 – 147.50	-4.50	3.40	0.96	0.05 – 0.99
23rd	11:00 – 11:30	147.50 – 149.38	-1.88	1.45	0.99	0.38 – 0.99
24th	11:30 – 12:00	149.38 – 150.88	-1.50	1.67	0.98	0.92 – 0.99

Table – 3.4 Intra-session reliability of Zephyr Bioharness and Fitbit Charge heart rate variables throughout Recovery.

Measure	Time (min: sec)	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	151.50–132.47	19.03	10.47	0.44	-0.02 – 0.73
2nd	0:30 – 1.00	132.47–113.00	19.47	10.10	0.44	-0.06 – 0.77
3rd	1:00 – 1:30	113.02-103.20	9.82	4.22	0.71	-0.02 – 0.89
4th	1:30 – 2:00	103.28 – 97.30	5.98	5.63	0.82	0.27 – 0.93
5th	2:00 – 2:30	97.30 – 95.50	1.80	3.72	0.91	0.84 – 0.95
6th	2:30 – 3:00	95.57 – 91.30	4.27	4.44	0.86	0.48 – 0.94
7th	3:00 – 3:30	91.30 – 90.00	1.30	2.65	0.94	-0.89 – 0.96
8th	3:30 – 4:00	90.07 – 89.90	0.17	3.21	0.91	0.85 – 0.95
9th	4:00 – 4:30	89.95 – 86.90	3.05	3.88	0.88	0.70 – 0.94
10th	4:30 – 5:00	86.90 – 86.90	0	2.59	0.95	0.92 – 0.97
11th	5:00 – 5:30	86.93 – 86.20	0.73	3.05	0.93	0.88 – 0.95
12th	5:30 – 6:00	86.27 – 86.00	0.27	2.74	0.94	0.90 – 0.96
13th	6:00 – 6:30	86.07 – 85.00	1.07	2.47	0.96	0.93 – 0.97
14th	6:30 – 7:00	85.05 – 85.80	-0.75	2.64	0.95	0.91 – 0.97
15th	7:00 – 7:30	85.82 – 84.60	1.22	1.94	0.97	0.94 – 0.98
16th	7:30 – 8:00	84.63 – 84.70	-0.07	1.26	0.98	0.98 – 0.99
17th	8:00 – 8:30	84.75 – 83.90	0.85	1.45	0.98	0.97 – 0.99
18th	8:30 – 9:00	83.97 – 83.80	0.17	1.74	0.97	0.95 – 0.98
19th	9:00 – 9:30	83.83 – 82.60	1.23	1.94	0.97	0.93 – 0.98
20th	9:30 – 10:00	82.60 – 82.70	-0.10	1.56	0.98	0.96 – 0.98

Table – 3.4 Intra-session reliability of Zephyr and Fitbit Charge heart rate variables throughout Recovery.

Measure	Time (min: sec)	FitBit – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0.30	150.88 – 131.27	19.61	11.00	0.50	-0.10 – 0.80
2nd	0:30 – 1.00	131.27 – 113.78	17.49	11.90	0.45	-0.09 – 0.76
3rd	1:00 – 1:30	113.78 – 103.82	9.96	8.22	0.70	0.01 – 0.89
4th	1:30 – 2:00	103.82 – 98.73	5.09	4.66	0.85	0.43 – 0.94
5th	2:00 – 2:30	98.73 – 94.97	3.76	4.48	0.86	0.62 – 0.94
6th	2:30 – 3:00	94.97 – 90.07	4.90	4.38	0.85	0.24 – 0.95
7th	3:00 – 3:30	90.07 – 88.40	1.67	2.70	0.94	0.88 – 0.97
8th	3:30 – 4:00	88.40 – 87.37	1.03	3.05	0.91	0.86 – 0.95
9th	4:00 – 4:30	87.37 – 85.90	1.47	2.73	0.93	0.87 – 0.96
10th	4:30 – 5:00	85.90 – 85.93	-0.03	2.30	0.95	0.92 – 0.97
11th	5:00 – 5:30	85.93 – 84.92	1.01	2.65	0.94	0.89 – 0.96
12th	5:30 – 6:00	84.92 – 83.55	1.37	2.89	0.92	0.87 – 0.96
13th	6:00 – 6:30	83.55 – 84.33	-0.78	3.32	0.90	0.83 – 0.94
14th	6:30 – 7:00	84.33 – 83.73	0.60	2.64	0.93	0.88 – 0.96
15th	7:00 – 7:30	83.73 – 83.20	0.53	1.47	0.98	0.96 – 0.99
16th	7:30 – 8:00	83.20 – 83.52	-0.32	1.72	0.97	0.95 – 0.98
17th	8:00 – 8:30	83.52 – 83.05	0.47	1.70	0.97	0.95 – 0.98
18th	8:30 – 9:00	83.05 – 82.87	0.18	1.92	0.96	0.94 – 0.98
19th	9:00 – 9:30	82.87 – 81.45	1.42	2.37	0.94	0.87 – 0.97
20th	9:30 – 10:00	81.45 – 81.33	0.12	1.15	0.98	0.97 – 0.98

Table – 3.5 Inter-session reliability of Zephyr Bioharness *and Fitbit Charge* heart rate variables at Rest.

Measure	Time (min: sec) 1st 2nd	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	71.32– 71.53	-0.21	1.73	0.94	0.91 – 0.96
2nd	1:00 – 1:00	71.62– 71.78	-0.16	1.90	0.93	0.89 – 0.96
3rd	1:30 – 1:30	71.67– 71.50	0.17	2.10	0.92	0.87 – 0.95
4th	2:00 – 2:00	71.72– 71.87	-0.15	2.26	0.91	0.85 – 0.94
5th	2:30 – 2:30	71.30– 71.55	-0.25	2.01	0.91	0.86 – 0.95
6th	3:00 – 3:00	71.57– 71.28	0.29	2.26	0.90	0.85 – 0.94
7th	3:30 – 3:30	71.62– 71.37	0.25	2.17	0.91	0.86 – 0.94
8th	4:00 – 4:00	71.50– 71.70	-0.20	2.17	0.92	0.86 – 0.94
9th	4:30 – 4:30	71.73– 71.25	0.48	2.28	0.91	0.85 – 0.94
10th	5:00 – 5:00	71.35– 71.57	-0.22	2.18	0.91	0.86 – 0.94
11th	5:30 – 5:30	71.63– 71.57	0.06	2.31	0.90	0.84 – 0.94
12th	6:00 – 6:00	71.67– 71.78	-0.11	2.30	0.91	0.84 – 0.94
13th	6:30 – 6:30	71.32– 71.23	0.09	1.85	0.93	0.90 – 0.96
14th	7:00 – 7:00	71.48– 71.68	-0.20	1.79	0.94	0.91 – 0.96
15th	7:30 – 7:30	71.35– 71.60	-0.25	1.94	0.92	0.87 – 0.95
16th	8:00 – 8:00	71.58– 72.00	-0.42	2.24	0.90	0.84 – 0.94
17th	8:30 – 8:30	71.85– 71.83	0.02	2.13	0.92	0.86 – 0.95
18th	9:00 – 9:00	71.80– 71.73	0.07	2.37	0.90	0.83 – 0.93
19th	9:30 – 9:30	71.55– 71.87	-0.32	2.10	0.92	0.87 – 0.95
20th	10:00 – 10:00	71.55– 71.85	-0.30	2.14	0.91	0.86 – 0.94

Table – 3.5 Inter-session reliability of Zephyr Bioharness and Fitbit heart rate variables at Rest.

Measure	Time (min: sec) 1st 2nd	FitBit – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	71.92 – 71.87	0.05	1.98	0.93	0.90 – 0.96
2nd	1:00 – 1:00	72.00 – 72.32	-0.32	2.02	0.93	0.88 – 0.95
3rd	1:30 – 1:30	71.72 – 71.72	0.00	2.17	0.91	0.86 – 0.94
4th	2:00 – 2:00	71.90 – 72.05	-0.15	1.94	0.93	0.88 – 0.95
5th	2:30 – 2:30	71.33 – 71.87	-0.54	2.01	0.92	0.87 – 0.95
6th	3:00 – 3:00	71.53 – 71.65	-0.12	2.54	0.88	0.80 – 0.92
7th	3:30 – 3:30	71.78 – 71.66	0.12	2.27	0.90	0.84 – 0.94
8th	4:00 – 4:00	71.95 – 72.08	-0.13	2.34	0.90	0.83 – 0.93
9th	4:30 – 4:30	71.87 – 71.43	0.44	2.43	0.89	0.82 – 0.93
10th	5:00 – 5:00	71.66 – 72.02	-0.36	2.26	0.91	0.84 – 0.94
11th	5:30 – 5:30	71.86 – 71.77	0.09	2.60	0.88	0.80 – 0.92
12th	6:00 – 6:00	71.65 – 72.05	-0.40	2.20	0.91	0.85 – 0.94
13th	6:30 – 6:30	71.81 – 71.66	0.15	2.04	0.92	0.87 – 0.95
14th	7:00 – 7:00	71.63 – 71.82	-0.19	1.81	0.94	0.90 – 0.96
15th	7:30 – 7:30	71.63 – 71.87	-0.24	2.67	0.92	0.87 – 0.95
16th	8:00 – 8:00	71.85 – 72.30	-0.45	2.35	0.89	0.82 – 0.93
17th	8:30 – 8:30	71.91 – 72.33	-0.42	2.28	0.91	0.84 – 0.94
18th	9:00 – 9:00	72.02 – 72.15	-0.13	2.30	0.91	0.84 – 0.94
19th	9:30 – 9:30	72.00 – 72.30	-0.30	1.83	0.94	0.90 – 0.96
20th	10:00 – 10:00	71.55 – 72.30	-0.75	2.38	0.89	0.82 – 0.93

Table – 3.6 Inter-session reliability of Zephyr Bioharness *and Fitbit Charge* heart rate variables during mCAFT.

Measure	Time (min: sec) 1st 2nd	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	102.93 – 99.78	3.15	3.15	0.93	0.63 – 0.97
2nd	1:00 – 1:00	111.65–108.40	3.25	3.70	0.93	0.75 – 0.97
3rd	1:30 – 1:30	118.42–115.48	2.94	3.65	0.94	0.84 – 0.97
4th	2:00 – 2:00	120.87–117.70	3.17	3.44	0.96	0.82 – 0.98
5th	2:30 – 2:30	122.95–120.17	2.78	3.23	0.97	0.88 – 0.98
6th	3:00 – 3:00	124.65–122.60	2.05	3.12	0.97	0.94 – 0.98
7th	3:30 – 3:30	114.88–113.04	1.84	3.40	0.95	0.90 – 0.97
8th	4:00 – 4:00	124.34–121.36	2.98	3.93	0.93	0.83 – 0.97
9th	4:30 – 4:30	128.46–125.73	2.73	3.46	0.95	0.86 – 0.98
10th	5:00 – 5:00	130.32–128.31	2.01	3.22	0.96	0.92 – 0.98
11th	5:30 – 5:30	132.78–131.00	1.78	3.33	0.96	0.93 – 0.98
12th	6:00 – 6:00	134.35–132.47	1.88	3.20	0.97	0.94 – 0.98
13th	6:30 – 6:30	124.51–122.49	2.02	3.26	0.93	0.85 – 0.96
14th	7:00 – 7:00	131.80–129.84	1.96	3.16	0.93	0.85 – 0.96
15th	7:30 – 7:30	138.08–135.59	2.49	3.15	0.93	0.79 – 0.97
16th	8:00 – 8:00	141.87–138.64	3.23	3.43	0.92	0.65 – 0.97
17th	8:30 – 8:30	143.92–141.28	2.64	3.21	0.93	0.79 – 0.97
18th	9:00 – 9:00	146.62–143.69	2.93	3.34	0.93	0.74 – 0.97
19th	9:30 – 9:30	128.63–126.13	2.50	3.28	0.94	0.72 – 0.98
20th	10:00 – 10:00	139.00–135.38	3.62	4.50	0.92	0.62 – 0.98
21th	10:30 – 10:30	145.38–140.75	4.63	4.64	0.91	0.46 – 0.98
22nd	11:00 – 11:00	148.25–145.00	3.25	3.77	0.94	0.68 – 0.98
23rd	11:30 – 11:30	149.88–146.13	3.75	4.17	0.93	0.62 – 0.98
24th	12:00 – 12:00	151.50–147.38	4.12	3.88	0.94	0.49 – 0.99

Table – 3.6 Inter-session reliability of Zephyr Bioharness and Fitbit Charge heart rate variables during mCAFT.

		FitBit – Heart Rate				
Measure	Time (min: sec) 1st 2nd	M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	89.42 – 87.89	1.53	3.85	0.91	0.84 – 0.94
2nd	1:00 – 1:00	103.93 – 101.57	2.36	4.04	0.91	0.84 – 0.95
3rd	1:30 – 1:30	114.63 – 112.08	2.55	3.70	0.95	0.88 – 0.97
4th	2:00 – 2:00	119.45 – 116.72	2.73	4.40	0.93	0.87 – 0.96
5th	2:30 – 2:30	121.20 – 119.12	2.08	3.70	0.96	0.92 – 0.97
6th	3:00 – 3:00	122.63 – 120.68	1.95	3.53	0.97	0.94 – 0.98
7th	3:30 – 3:30	110.14 – 109.16	0.98	3.53	0.95	0.91 – 0.97
8th	4:00 – 4:00	118.78 – 116.47	2.31	3.76	0.95	0.89 – 0.97
9th	4:30 – 4:30	124.22 – 121.55	2.67	4.24	0.95	0.88 – 0.97
10th	5:00 – 5:00	127.60 – 125.10	2.50	3.60	0.97	0.91 – 0.98
11th	5:30 – 5:30	129.83 – 127.73	2.10	3.50	0.96	0.93 – 0.98
12th	6:00 – 6:00	132.67 – 130.27	2.40	4.25	0.95	0.91 – 0.97
13th	6:30 – 6:30	120.87 – 117.87	3.00	3.70	0.92	0.76 – 0.96
14th	7:00 – 7:00	127.36 – 123.90	3.46	4.27	0.91	0.72 – 0.96
15th	7:30 – 7:30	134.05 – 131.62	2.43	3.65	0.92	0.83 – 0.96
16th	8:00 – 8:00	138.56 – 135.92	2.64	3.62	0.92	0.81 – 0.96
17th	8:30 – 8:30	141.50 – 139.33	2.17	4.73	0.89	0.78 – 0.94
18th	9:00 – 9:00	144.00 – 140.77	3.23	3.28	0.93	0.65 – 0.97
19th	9:30 – 9:30	124.87 – 121.25	3.62	3.93	0.90	0.48 – 0.97
20th	10:00 – 10:00	135.38 – 129.38	6.00	4.68	0.85	0.20 – 0.97
21th	10:30 – 10:30	143.00 – 137.50	5.50	4.70	0.88	0.33 – 0.97
22nd	11:00 – 11:00	147.50 – 142.13	5.37	4.77	0.90	0.24 – 0.98
23rd	11:30 – 11:30	149.38 – 143.88	5.50	4.88	0.89	0.26 – 0.98
24th	12:00 – 12:00	150.88 – 146.00	4.88	4.51	0.92	0.35 – 0.98

Table – 3.7 Inter-session reliability of Zephyr Bioharness *and Fitbit Charge* heart rate variables throughout Recovery.

Measure	Time (min: sec) 1st 2nd	Zephyr – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	132.40–129.65	2.75	4.10	0.90	0.78 – 0.95
2nd	1:00 – 1:00	113.02–114.05	-1.03	3.85	0.93	0.89 – 0.96
3rd	1:30 – 1:30	103.28–105.38	-2.10	4.12	0.91	0.84 – 0.95
4th	2:00 – 2:00	97.30 – 99.77	-2.47	4.66	0.87	0.76 – 0.92
5th	2:30 – 2:30	95.57 – 97.10	-1.53	4.37	0.87	0.79 – 0.92
6th	3:00 – 3:00	91.30 – 93.07	-1.77	4.14	0.88	0.79 – 0.93
7th	3:30 – 3:30	90.07 – 92.63	-2.56	3.74	0.87	0.73 – 0.93
8th	4:00 – 4:00	89.95 – 91.58	-1.63	3.75	0.88	0.79 – 0.93
9th	4:30 – 4:30	86.90 – 88.93	-2.03	3.91	0.89	0.80 – 0.93
10th	5:00 – 5:00	86.93 – 88.13	-1.20	3.31	0.91	0.86 – 0.95
11th	5:30 – 5:30	86.27 – 87.57	-1.30	3.59	0.88	0.80 – 0.93
12th	6:00 – 6:00	86.07 – 87.22	-1.15	3.15	0.92	0.87 – 0.96
13th	6:30 – 6:30	85.05 – 87.13	-2.08	2.95	0.93	0.84 – 0.97
14th	7:00 – 7:00	85.82 – 86.70	-0.88	2.65	0.94	0.90 – 0.96
15th	7:30 – 7:30	84.63 – 86.00	-1.37	3.15	0.93	0.87 – 0.96
16th	8:00 – 8:00	84.75 – 85.20	-0.45	2.84	0.93	0.89 – 0.96
17th	8:30 – 8:30	83.97 – 85.15	-1.18	2.78	0.93	0.88 – 0.96
18th	9:00 – 9:00	83.83 – 84.45	-0.62	2.95	0.91	0.86 – 0.95
19th	9:30 – 9:30	82.60 – 83.97	-1.37	3.10	0.91	0.84 – 0.94
20th	10:00 – 10:00	82.72 – 84.40	-1.68	3.11	0.89	0.80 – 0.94

Table – 3.7 Inter-session reliability of Zephyr Bioharness and Fitbit heart rate variables throughout Recovery.

Measure	Time (min: sec) 1st 2nd	FitBit – Heart Rate				
		M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0:00 – 0:00	131.27 – 131.82	-0.55	4.67	0.80	0.63 – 0.85
2nd	1:00 – 1:00	113.78 – 114.47	-0.69	4.15	0.93	0.88 – 0.96
3rd	1:30 – 1:30	103.82 – 106.18	-2.36	4.50	0.91	0.83 – 0.95
4th	2:00 – 2:00	98.73 – 99.82	-1.09	4.45	0.84	0.74 – 0.90
5th	2:30 – 2:30	94.97 – 95.02	-0.05	4.56	0.83	0.73 – 0.90
6th	3:00 – 3:00	90.07 – 92.63	-2.56	4.20	0.89	0.78 – 0.94
7th	3:30 – 3:30	88.40 – 90.61	-2.21	4.30	0.87	0.77 – 0.92
8th	4:00 – 4:00	87.37 – 89.25	-1.88	3.73	0.89	0.80 – 0.94
9th	4:30 – 4:30	85.90 – 87.98	-2.08	4.20	0.85	0.75 – 0.91
10th	5:00 – 5:00	85.93 – 87.45	-1.52	3.63	0.89	0.82 – 0.94
11th	5:30 – 5:30	84.92 – 86.18	-1.26	3.30	0.91	0.85 – 0.94
12th	6:00 – 6:00	83.55 – 86.03	-2.48	3.90	0.87	0.75 – 0.93
13th	6:30 – 6:30	84.33 – 85.60	-1.27	3.62	0.88	0.81 – 0.93
14th	7:00 – 7:00	83.73 – 85.43	-1.7	3.10	0.90	0.82 – 0.95
15th	7:30 – 7:30	83.20 – 84.83	-1.63	2.90	0.91	0.84 – 0.95
16th	8:00 – 8:00	83.51 – 84.77	-1.26	2.98	0.91	0.85 – 0.95
17th	8:30 – 8:30	83.05 – 85.03	-1.98	3.17	0.90	0.79 – 0.94
18th	9:00 – 9:00	82.87 – 84.32	-1.45	3.38	0.87	0.79 – 0.92
19th	9:30 – 9:30	81.45 – 83.43	-1.98	3.35	0.86	0.74 – 0.92
20th	10:00 – 10:00	81.33 – 83.07	-1.74	3.35	0.82	0.70 – 0.89

Table – 3.8 Test-retest reliability of Fitbit variables Steps Taken and Energy Expenditure during mCAFT.

Variable		FitBit – Steps Taken and Energy Expenditure			
	M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
Steps Taken	899.00 – 893.00	6.00	17.00	0.99	0.98 – 0.99
Energy Expenditure	131.00 – 129.00	2.00	9.00	0.93	0.89 – 0.96

Table – 3.9 Combined Inter-session reliability of FC-HR heart rate variable.

Measures at Rest	FitBit – Heart Rate					
	Time (min: sec)	M_1 and M_2	Mean Diff.	SEM	ICC	Confidence Interval (95%)
1st	0.00 - 0.30	71.92 – 71.86	0.06	1.98	0.94	0.90 – 0.96
1st - 3rd	0.00 – 1:30	71.88 – 71.97	-0.09	2.00	0.95	0.92 – 0.97
1st - 5th	0.00 – 2:30	71.77 – 71.96	-0.19	1.99	0.96	0.93 – 0.98
1st - 20th	0.00 – 10:00	71.78 – 71.96	-0.18	2.09	0.95	0.93 – 0.97
Measures at mCAFT						
1st	0.00 - 0.30	89.42 – 87.97	1.45	3.85	0.91	0.84 – 0.94
1st - 3rd	0.00 – 1:30	102.66 - 100.54	2.12	3.79	0.94	0.87 – 0.97
1st - 5th	0.00 – 2:30	109.73 – 107.49	2.24	3.89	0.95	0.89 – 0.98
1st - 24th	0.00 – 10:00	129.63 – 128.72	0.91	3.98	0.85	-0.81 – 0.99
Measures at Recovery						
1st	0.00 - 0.30	131.27 – 131.83	-0.56	4.67	0.80	0.63 – 0.85
1st - 3rd	0.00 – 1:30	116.29 – 117.49	-1.2	4.44	0.94	0.89 – 0.96
1st - 5th	0.00 – 2:30	108.51 – 109.46	-0.95	4.46	0.93	0.89 – 0.96
1st - 20th	0.00 – 10:00	90.61 – 92.20	-1.59	4.10	0.90	0.86 – 0.96

CHAPTER 4:

Title of Manuscript:

Validity and Inter-Instrument Agreement of Fitbit Charge Measures of Heart Rate and Activity at Rest, During The Modified Canadian Aerobic Fitness Test and Recovery.

Authors:

Goris Nazari PT, MSc (C) [a],

Joy C MacDermid PT, PhD [b-c],

Kathryn E Sinden R. Kin., PhD [d],

Julie Richardson PT, PhD [a],

Ada Tang PT, PhD [a],

[a] McMaster University, School of Rehabilitation Science, 1400 Main Street West, Hamilton, Ontario, Canada.

[b] Physical Therapy, Western University, London, Ontario, Canada.

[c] Roth McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London, Ontario, Canada

[d] Department of Kinesiology and Physical Education, McGill University, 475 Pine Street, Montreal, Québec, Canada.

Corresponding Author:

Goris Nazari, McMaster University, School of Rehabilitation Science, 1400 Main Street West IAHS 403, Hamilton, Ontario, Canada. E-mail: nazarigs@mcmaster.ca

Acknowledgements:

Joy C MacDermid was supported by a Canadian Institutes of Health Research (CIHR) Chair in Gender, Work and Health and the Dr James Roth Chair in Musculoskeletal Measurement and Knowledge Translation. This work was supported by an operating grant from the Ministry of Labour - Grant Number #13-R-027. Ada Tang was supported by a personnel award from the Heart and Stroke Foundation, Ontario Provincial Office (CS I 7468).

Ethics Approval:

This study was approved by Hamilton Integrated Research Ethics Board (HiREB) of McMaster University, Hamilton, Ontario, Canada**.**

Conflict of Interest:

There are no financial gains or conflict of interest with the content of this research article.

.

Abstract

Introduction: Advances in technology has assisted in development of Wearable Physiological Monitoring devices that are small, non-invasive as well as easy to use in capturing and monitoring of various physiological measures. Therefore, the purpose of this study was to establish the validity and inter-device agreement between the Fitbit Charge and Zephyr Bioharness devices at Rest, during a sub-maximal test and throughout Recovery.

Methods: Sixty participants were recruited (30 females, 48 ± 15 years) and (30 males, 48 ± 15 years) using convenience and snowball sampling approaches from McMaster University. The performance of Zephyr and Fitbit devices were assessed throughout three phases; rest, Modified Canadian Aerobic Fitness Test and recovery. To establish concurrent validity, the Fitbit device variables; heart rate, steps taken and energy expenditures, were compared with two adopted criterion measures; the Zephyr Bioharness for heart rate, and Fitbit-One for steps taken and energy expenditure variables.

Results: Pearson's correlation coefficients (Mean differences) for heart rate variable were 0.97 – 0.99 (-0.60 – 0.02) at rest, 0.89 – 0.99 (13.51 – 0.62) at submaximal testing and 0.70 – 0.84 (-0.54 – 2.52) throughout recovery. In addition, for steps taken and energy expenditure variables Pearson's correlations (Mean differences) were 0.98 (79.43) and 0.80 (39.13) respectively. The average agreement bias of heart rate in pair-wise device comparison indicated mean differences of -0.20, 4.00 and 1.00 at rest, sub-maximal testing and recovery respectively. Steps taken and energy expenditure comparison yielded larger average biases of 79.40 and 39.20 respectively.

Conclusion: The FC-HR device heart rate variable demonstrated strong to very strong correlations when concurrently compared with ZB, and provided valuable information regarding its interchangeable use in a sample of sixty healthy male and female participants of various age groups at rest, during a standardized submaximal fitness and during recovery phases following the test.

Introduction

Study Background:

Advances in technology has promoted the development of Wearable Physiological Monitoring (WPM) devices that are small, non-invasive as well as easy to use in capturing and monitoring physiological measures across various fields including personnel in the fire service, or construction workers as well as promoting changes in physical activity levels (Johnstone et al., 2012b; Gatti et al., 2014; Smith et al., 2014; Tully et al., 2014). These advances in technology have led to a constant development of newer WPM devices. One such a device is the Fitbit Charge Heart Rate (FC-HR) a wrist band capable of recording continuous wrist-based heart rate and other physical activity components (Fitbit Charge HR Manual 2015).

Heart rate monitoring can have important implications. The maximum heart rate (HR_{max}) is the highest heart rate an individual can achieve without severe problems (Atwal, Porter & MacDonald 2002). An individual's maximum heart rate is age dependent and generally decreases with age. Exceeding maximum heart rate can be dangerous to your health, as exertion at extreme intensities are associated with increased risks for cardiac events (Atwal, Porter & MacDonald 2002). Heart rate monitoring is among one of the most common techniques used to determine work related physiological demands (Gatti et al., 2014). Excessive work related physiological demands can jeopardise the safety and productivity as it lowers the worker's attentiveness, motivation as well as the capacity to perform muscular work (Gatti et al., 2014). In addition, in athletes, heart rate monitoring provides valuable information as during early stages of overtraining the maximal heart rates as well

as submaximal heart rates may be decreased, and while resting and sleeping heart rates may be increased (Jeukendrup & Van Dieme 1998).

In research, standardised methods to measure physical activity have been developed which involve the use of accelerometers (Lee, Kim & Welk 2014). Physical inactivity is a causal factor to the ever expanding healthcare costs associated with obesity and chronic disease (Colditz 1999; Physical Activity Guidelines Advisory Committee Report 2008). Methods of prevention and treatment for obesity and chronic disease involve increasing physical activity levels by quantifying the number of steps taken and total energy expenditures (Takacs et al., 2014).

Both reliability and validity measures of a device are important for its utilization (Streiner & Norman 1995). In our previous study (Chapter – 3), FC-HR device demonstrated excellent intra-session reliability (stability) and inter-session (test-retest) reliability measures. Validity is refereed to the degree to which a device measures what it is intended to measure (Brazier & Deverill 1999). Concurrent validity is established when simultaneous recordings are made by a device (to be validated) and its criterion (Portney & Watkins 1993). However, the validity between two devices does not warrant agreement, as a lack of agreement in spite of high correlations can be indicative of presence of bias (Bunce 2009). Since conceptual differences exist between validity and agreement parameters, and that there is a paucity of reports in the current literature about measures related to FC-HR device, therefore we aimed to establish the validity and agreement parameters of FC-HR, against two reliable and valid devices; Zephyr Bioharness (ZB) and

Fitbit-ONE (F-ONE), during three phases; at rest, during a sub-maximal test and throughout recovery following the exercise test.

Study Research Questions:

The overall aim of this study was to determine the validity and levels of agreement between the ZB and FC-HR devices. Specifically; 1) To determine the concurrent validity and the levels of agreement between ZB and FC-HR heart rate variables at rest, during mCAFT and throughout recovery. 2) To establish the concurrent validity and inter-device agreement between FC-HR and F-ONE number of steps taken and energy expenditure variables during Modified Canadian Aerobic Fitness Test.

Methods

Sampling and Recruitment:

Stratified convenience and snowballing sampling approaches were used and a total of sixty participants (30 females) were recruited from the School of Rehabilitation Science and Physical Activity Centre of Excellence, McMaster University. We received ethical approval for this study through the Hamilton Integrated Research Ethics Board (No. 0825),

Inclusion Criteria:

A Physical Activity Readiness Questionnaire (PAR-Q) was administered to both males and females between 20 – 69 years of age and only those with "No" answers to all seven PAR-Q questions were eligible to take part in the study (Canadian Society for Exercise

Physiology 1998). The ability to read, write and communicate in English was also a requirement.

Exclusion Criteria:

Individuals with "Yes" response/s to any of PAR-Q questions (Canadian Society for Exercise Physiology 1998).

Sample Size Calculations:

The sample size calculation was based upon our previous Zephyr and Fitbit Charge reliability study (Chapter – 3) with a null hypothesis test-retest reliability value of ICC 0.80 and the expectation of obtaining a test-retest reliability ICC of 0.90 (Donner & Eliasziw 1987; Johnstone et al., 2012; Johnstone et al., 2012b). Based on the calculations, a sample size of 60 participants was needed (Appendix A).

Wearable Physiological Monitoring devices

Fitbit Charge Heart Rate

Fitbit Charge Heart Rate (FC-HR) is a wristband that provides continuous automatic wrist-based heart rate through an optical heart rate monitor as well as number of steps taken and energy expenditure by a 3-axis accelerometry sensor (Fitbit Charge HR Manual 2015). The device synchronises with personal computers, is powered by a Lithium-polymer battery, and lasts for 5-days when charged for two hours and stores data up to 30 days. Based on our previous study, FC-HR device demonstrated excellent reliability measures.

Zephyr BioHarness

The Zephyr BioHarness (ZB) consists of an adjustable chest belt featuring conductive fabric sensors and an electronic BioModule that snaps on to the belt. The device fits comfortably on the chest at the lower sternum for both men and women and weighs 85 grams. The device is monitors and records physiological measures such as; heart rate, respiratory rate and estimated core temperature, posture and activity levels (Zephyr Technology Manual 2012). Recording of cardiac electrical impulses by conductive fabric skin electrodes on the wearer enables the monitoring and measure of the heart rate. The cardiac impulses are then sent to and processed by the BioModule as beats per minute ($b \cdot min^{-1}$) (Zephyr Technology Manual 2012). With three-hour charging of the BioModule, 26 hours of data is recorded and can log up to 20 days of data. Based on previous studies, ZB is considered as a reliable and valid device in assessing heart rate measures, and its agreement parameters against gold standard measures have also been established, suggestive of interchangeable use. (Johnstone et al., 2012; Johnstone et al., 2012a; Johnstone et al., 2012b; Kim et al., 2013; Gatti et al., 2014; Smith et al., 2014; Dolezal et al., 2014; Flanagan et al., 2014; Rawstorn et al., 2015).

Fitbit ONE

The Fitbit ONE (F-ONE) is a small (48.0 x 19.3 x 9.6 mm) and lightweight (8g) advanced tri-axial accelerometry-based device which can be worn on the hip, in the front pocket of pants or shorts and it tracks physical activity and measures sleep quality (Fitbit-One Manual 2012). Its physical activity recording features include; number of steps taken, energy

expenditure, floors climbed and distance travelled. It is powered by a Lithium-ion polymer battery, stores data for up to 23-days and the captured data can be uploaded to a personal computer (Fitbit-One Manual 2012). The reliability and validity of F-ONE has been reported in the literature (Lee, Kim & Welk 2014; Takacs et al., 2014; Diaz et al. 2015; Ferguson et al., 2015).

Modified Canadian Aerobic Fitness Test:

The Modified Canadian Aerobic Fitness Test (mCAFT) is a multistage submaximal step test that consists of eight-stages (Canadian Society for Exercise Physiology 1998).The initial stepping stage and 85% of maximum heart rate were calculated based on each participant's age and gender. The 85% maximum heart rate value is referred to as the ceiling post-exercise heart rate (Canadian Society for Exercise Physiology 1998). During a stepping session, the participants completed a three-minute stepping on a double 20.3 cm step stool at a predetermined cadence (foot-plants/minutes) corresponding to the assigned stepping stage. At the end of each three-minute stepping session, participants' heart rate was measured and compared with the predetermined ceiling post-exercise heart rate, and if the measured heart rate did not equal or exceeded the predetermined heart rate value the participants proceeded to the next three-minute stepping session (Canadian Society for Exercise Physiology 1998). The participants performed these progressively demanding three-minute stepping sessions until they achieved a heart rate that equated or exceeded the ceiling post-exercise heart rate (Canadian Society for Exercise Physiology 1998). Once the predetermined heart rate was achieved, the test was completed. The mCAFT sub-maximal

test has been identified as a reliable, valid, standardized and feasible, it provides individual's with an estimated levels of maximal oxygen uptake ($VO_{2\text{-}max}$) and involves a functional task – stair climbing (Canadian Society for Exercise Physiology 1998; Weller et al., 1994; Weller & Corey 1998)

Concurrent Validity of FC-HR device:

To establish its concurrent validity, the FC-HR variables; Heart Rate, Steps Taken and Energy Expenditures, were concurrently compared with two adopted criterion measures. The criterion measures within this study were the ZB for heart rate variable and F-ONE for steps taken and energy expenditure variables.

Study Design:

Standardised procedures were followed where the participants were emailed the mCAFT Preliminary Participant Instructions Form (Do's and Don'ts prior to their session – Appendix E) 48 hours prior to their visit. At their first visit, the principal investigator explained the study to the participants, administered the PAR-Q as well as obtained a written signed Consent Form. Next, the participants' resting heart rate, resting blood pressure, height (m) and body weight (kg) measures, were recorded[16]. Then, the devices; ZB, FC-HR and the F-ONE, were fitted. Participants were then required to be seated for a period of ten minutes which was considered – rest period. Following this ten-minute "rest" period, the mCAFT – "activity phase" was administered. The mCAFT termination was

followed by a ten-minute "recovery" period where the participants were required to be seated again

Measurement Protocol:

To determine the concurrent validity of FC-HR and establish its levels of agreement with ZB and F-ONE devices, participants' heart rate were recorded using the FC-HR and ZB devices during three phases. At rest, a total of twenty measures for ten minutes, during the mCAFT a total of twenty-four measures for twelve minutes and throughout recovery a total of twenty measures for ten minutes, all at 30-second consecutive intervals were recorded. In addition, the FC-HR and F-ONE devices were both used to record the total number of steps taken and energy expenditures during the mCAFT.

Statistical Analysis

Demographic characteristics of the sample stratified by gender; women and men, including age, height, weight, body mass index, age-related heart rate maximum were described using means, standard deviations, minimum and maximum scores.

The ZB data file was exported into Microsoft Excel 2016, and was stratified into 30-second time intervals. We collected FC-HR heart rate data through the Fitbit Dashboard software using The Fitbit App, at 30-second time intervals. In addition, FC-HR steps taken and energy expenditure data were extracted using the recommended software suggested by the Fitbit manufacturers. To indicate concurrent validity, Pearson correlation coefficients (r)

were used to compare 1) strength of association between the ZB and FC-HR heart rate variables at rest, during mCAFT and throughout recovery phases, and 2) strength of association between the FC-HR and F-ONE steps taken and energy expenditure variables during the mCAFT. Two hypotheses were tested; 1) ZB vs. FC-HR heart rate measures at all the three phases would demonstrate strong correlations, and 2) FC-HR vs. F-ONE steps taken and energy expenditure recordings during mCAFT would similarly display strong correlations. The absolute value for the strength of correlation (r) using the guide suggested by (Evans 1996) were; 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong", 0.80-1.00 "very strong"[35]. To determine the levels of agreement between ZB vs. FC-HR as well as FC-HR vs. F-ONE, MedCalc software bvba, version 16.2.1 was used to calculate Bland and Altman plots of individual differences against the mean of the two measures for all the three phases (Bland & Altman 1986). Individual agreement between each two devices was then summarised by the mean difference, the 95% limits of agreement (± 1.96 times standard deviations). To examine average agreement/ bias between devices, the mean differences were tested using a one-sample t-test and mean differences, standard error of differences, p - values and 95% Confidence intervals reported. Analyses were performed using IBM SPSS Statistics software version 22.0 and a significance level of $p \leq 0.05$ was considered statistically significant.

Results

Sample

Sixty (30 females, 30 males) healthy participants were involved in this study. Table – 4.1 displays the demographic characteristics; age, height, weight, body mass index, age-related heart rate maximum, stratified by gender.

Concurrent Validity between ZB and FC-HR heart rate Variable:

Pearson's correlation coefficients and mean differences between ZB and FC-HR heart rate variables at rest ranged from 0.97 – 0.99 and -0.60 – 0.02 respectively (Table – 4.2). During the mCAFT, Pearson's correlations ranged from 0.89 – 0.99 with the largest mean difference of 13.51 and lowest mean difference of 0.62 between the ZB and FC-HR (Table – 4.3). The highest mean differences were noted during the first minute of each stage. At times [min : sec] 0:30 and 1:00 (mean differences; 13.51 and 7.72), times 3:30 and 4:00 (4.74 and 5.57), times 6:30 and7:00 (3.64 and 4.43), times 9:30 and 10:00 (3.75 and 3.62) were recorded. Throughout the recovery period, Pearson's correlations of 0.70 – 0.84 and Mean differences of -0.54 – 2.52 were reported between the ZB and FC-HR devices (Table – 4.4).

Concurrent Validity between FC-HR vs. F-ONE Steps Taken and Energy Expenditure variables:

During the mCAFT, Pearson's correlation coefficients between the FC-HR and F-ONE devices were 0.98 and 0.80 with mean differences of 79.43 and 39.13 for the steps taken and energy expenditure variables respectively (Table – 4.5). The Pearson's correlation coefficients for these variables were not assessed at rest or throughout recovery.

Inter-Device Levels of Agreement:

The average agreement bias of heart rate in pair-wise device comparison indicated mean differences (95% Confidence Intervals) of -0.20 (-0.10 – -0.30), 4.00 (3.70 – 4.30) and 1.00 (0.55 – 1.45) at rest, during mCAFT and during recovery respectively (Table – 4.6). In addition, the steps taken and energy expenditure comparisons yielded mean differences (95% Confidence Intervals) of 79.40 (53.70 – 105.10) and 39.20 (34.00 – 44.40) respectively (Table – 4.6). However, when assessing heart rate individual levels of agreement, the Bland and Altman plots displayed wider (95% Limits of Agreement) for the mCAFT (48.40 – -40.30) and recovery (39.80 – -37.70) as compared to rest (5.20 – -5.7) (Figures 4.1A – E).

Discussion

This study established that strong to very strong correlations can be expected in healthy participants of different age groups when FC-HR device was concurrently compared with ZB devices to assess physiologic measure of heart rate at rest, during the mCAFT and throughout recovery. In addition, inter-instrument agreement measures of the devices demonstrated small average mean biases.

We used the submaximal fitness test as the stressor, to cause physiologic changes after a period of rest. The activity phase was followed by recovery, which was considered as resumption to a resting phase. With inclusion of different phases and how measurements

were taken across this study, we expect that transitioning through phases were considered as unstable intervals, which would explain the highest heart rate discrepancy during the first measure of submaximal testing. However, smaller mean heart rate difference between FC-HR and ZB were observed as the testing progressed.

Heart rate monitoring has important implications and the use of physical activity monitors to track and quantify number steps taken and cleric expenditures are common. Our study was the first to establish the validity measures of FC-HR, therefore on comparable studies were found. However, Pearson correlation coefficients and average mean differences for heart rate corresponded well with previously reported validity measures of ZB against electrocardiogram heart rate recordings at rest, during activity and recovery (Kim et al., 2013; Flanagan et al., 2014; Gatti et al., 2014; Smith et al., 2014; Dolezal et al., 2014). Pearson correlation coefficients for steps taken and energy expenditure in this study concurred well with Takacs et al., 2013 and Diaz et al., 2015 studies of investigating the validity measures of F-ONE steps taken and energy expenditures against gold standard techniques of direct observations and metabolic system/indirect calorimeter respectively. However, larger mean differences were reported in our study. And since in the two previous studies F-ONE has been compared against gold standard measures and smaller mean difference reported (Lee et al. 2014, Fergusson e al. 2015), it is reasonable to consider that in our study FC-HR has over-estimated the total number of steps taken and energy expenditures.

It is very rare to establish two identical results while measuring the same construct using two different devices (Bland & Altman 1986). Such measurements always involve some

degree of error. It is important to determine by how much the new device is likely to differ from the old one and whether the devices can be used interchangeably. According to Bland and Altman, computation of Pearson correlation coefficients to determine levels of agreement is inappropriate (Bland & Altman 1986). Pearson correlation coefficient assesses the linear relationship between two measures and is depended upon the range of measures; assessment of wider ranges often yields in higher correlations (Bunce 2009). Therefore, it is crucial to examine two aspects of agreement put forward by Bland and Altman. In average agreement estimation, a common t-test of mean differences of subjects against null hypothesis of no bias is conducted and 95% Confidence intervals reported as mean difference ± 1.96 × standard error of differences, whereas individual agreement calculations involve estimation of variability of differences and computation of 95% Limits of Agreement (LoA) as mean difference ± 1.96 × standard deviation of differences along with plotting of Bland-Altman plots (Bunce 2009). In our study, small mean differences and narrow confidence intervals indicated the presence of non-significant systematic differences between the two devices when recording heart rate measures. The wider limits of agreement during the mCAFT and subsequently throughout recovery could be due to large variability within our study sample and because of the nature of mCAFT. Since mCAFT is a sub-maximal test, it requires participants to achieve their 85% age-predicted maximum heart rates, therefore inclusion of participants ranging from 21 – 68 years of age for both men and women, and with calculated 85% heart rate maximum ranging from 129.00 – 169 beats/min. for men, and 129.00 – 167.00 beats/min. for women, could have contributed to these wider limits of agreement. However, following a period of proper

acclimatization, and rest phase likely considered as a stable condition, where participants did not have to achieve a specific age-related heart rate maximum, much narrower LoA were reported.

Strengths of this study included the fact that we established the validity and agreement parameters of FC-HR device during stable conditions and a standardised sub-maximal fitness test. In addition, we sampled a large number of participants and included a diverse age range and both genders. Despite this, limitations inherent in our approach and should be considered when interpreting our findings. We did not evaluate the performance of FC-HR against a gold standard (calibration). However, both the psychometric and agreement parameters of the criterion measures used in our study have been reported in the literature, and have been deemed as reliable and valid. We studied the validity and agreement parameters of FC-HR during human performance and did not assess the responsiveness of the device to accurately detect change when it has occurred. Future studies evaluating the reliability and validity properties of FC-HR device in assessing physiological measures, and longitudinal analysis to determine if the device will be responsive in clinical trails in different patient populations, are recommended.

Conclusions:

The FC-HR device heart rate variable demonstrated strong to very strong correlations when concurrently compared with ZB in a sample of sixty healthy male and female participants of various age groups during a resting, standardized submaximal fitness and

recovery phases. In addition, comparison of heart rate recordings between FC-HR and ZB throughout these phases provided valuable information and possible interchangeable use.

References

Atwal, S., Porter, J. and MacDonald, P. (2002). Cardiovascular effects of strenuous exercise in adult recreational hockey: The Hockey Heart Study. *CMAJ*, 166 (3), 303–307.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* **327** (8476), 307–10.

Brazier, J., & Deverill, M. (1999). A checklist for judging preference-based measures of health related qualityof life: learning from psychometrics. *Health Economics,* 8, 41–51.

Bunce, C. (2009). Correlation, Agreement, and Bland–Altman Analysis: Statistical Analysis of Method Comparison Studies. *Am J Ophthalmol,*148(1), 4-6.

Canadian Society for Exercise Physiology (1998). *The Canadian Physical Activity, Fitness and Lifestyle Appraisal* (2nd ed.). Ottawa, ON.

Colditz, G.A., (1999). Economic costs of obesity and inactivity. *Medicine & Science in Sports & Exercise*, 31, S663 – 667.

Diaz, K.M., Krupka, D.J., Chang, M.J., Peacock, J., Ma Y., Goldsmith, J., … Davidson, K.W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, 185, 138–140.

Dolezal, B.A., Boland D.M., Carney J., Abrazado M., Smith D.L. and Cooper C.B. (2014). Validation of heart rate derived from a physiological status monitor-embedded compression shirt against criterion ECG. *Journal of Occupational and Environmental Hygiene*, 11: 833–839.

Donner, A., and Eliasziw M. (1987). Sample Size Requirements for Reliability Studies. *Statistics in Medicine,* 6, 441-448.

Evans, J.D. (1996). *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Ferguson, T., Rowlands, A. V., Olds, T. and Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity,* 12 (42), 201-209.

Fitbit One Wireless Activity + Sleep Tracker (2012, September 17). *Fitbit One $_{TM}$ Product Manual*. Retrieved from https://www.fitbit.com/ca/one

Fitbit charge HR Wireless Heart Rate + Activity Wrist Band (2015, January*). Fitbit charge HR$_{TM}$ Product Manual Version 1.0.* Retrieved from https://www.fitbit.com/ca/chargehr.

Flanagan, S.D., Comstock, B.A., Dupont, W.H., Sterczala, A.R., Looney, D.P., Dombrowski, D.H., … Kraemer, W.J. (2014). Concurrent validity of the armour39 heart rate monitor strap. *Journal of Strength and Conditioning Research,* 28(3), 870–873.

Gatti, U.C., Schneider, S. and Migliaccio, G.C. (2014). Physiological condition monitoring of construction workers. *Automation in Construction* 44, 227–233.

Jeukendrup, A. & Van Dieme, A. (1998). Heart rate monitoring during training and competition in cyclist. *Journal of Sports Sciences*, 16, S91 - S99.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T., Mitchell A.C. and Garrett, A.T. (2012). Field based reliability and validity of the Bioharness^{TM} multivariable monitoring device. *Journal of Sports Science and Medicine* 11, 643-652.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012a). Bioharness^{TM} multivariable monitoring device: Part I: Validity. *Journal of Sports Science and Medicine* 11(3), 400- 408.

Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T. and Garrett, A.T. (2012b). Bioharness™ multivariable monitoring device: Part II: Reliability. *Journal of Sports Science and Medicine* 11(3), 409- 417.

Kim, J.H., Roberge, R., Powell, J. B., Shafer, A. B. and Williams, W. J. (2013). Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr Bioharness ™. *Int J Sports Med,* 34, 497–501.

Lee, J. M., Kim Y. and Welk G. J. (2014). Validity of consumer-based physical activity monitors. *Journal of the American College of Sports Medicine*, 14, 1840-1848.

Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Committee Report (2008). Washington, DC: U.S. Department of Health and Human Services.

Portney L.G., and Watkins M.P. (1993). Foundations of clinical research: applications to practice. Connecticut: Appleton & Lange.

Rawstorn, J.C., Gant, N., Warren, I., Doughty, R.N., Lever, N., Poppe, K.K. and Maddison, R. (2015). Measurement and data transmission: Validity of a multi-biosensor

system for real-time remote exercise monitoring among cardiac patients. *JMIR Rehabil Assist Technol,* 2 (1): e2.

Smith, D.L., Haller, J.M., Dolezal, B.A., Cooper, C.B. and Fehling, P.C. (2014). Evaluation of a wearable physiological status monitor during simulated fire fighting activities. *Journal of Occupational and Environmental Hygiene*, 11: 427–433.

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: a practical guide to their development and use* (2nded). New York: Oxford University Press.

Takacs, J., Pollock, C.L., Guenther, J.R., Bahar, M., Napier, C. and Hunt, M.A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17, 496–500

Tully, M. A., McBride, C., Heron, L. and Hunter, R. F. (2014). The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Research Notes*, 7, 952.

Weller, I.M. and Corey, P.N. (1998). A study of the reliability of the Canada Fitness Survey questionnaire. *Med Sci Sports Exerc,* 30, 1530-1536.

Weller, I.M., Thomas, S.G., Gledhill, N., Paterson, D. and Quinney, A. (1994). A study to validate the modified Canadian Aerobic Fitness Test. *Can J Appl Physiol*, 20, 211-221.

Zephyr Technology (2012, September 12). *BioHarness 3.0 User ManualBioHarness 3.0 User Manual*. Retrieved from http://zephyranywhere.com/products/bioharness-3/

Table – 4.1 Demographics

	Women (n = 30)				**Men (n = 30)**			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Age (yrs.)	48.00	15.00	23.00	68.00	48.00	15.00	21.00	68.00
Height (m)	1.70	0.05	1.60	1.77	1.78	0.06	1.68	1.88
Weight (kg)	69.00	11.00	50.00	100.00	79.00	8.50	64.00	97.00
Body Mass Index (kg/m^2)	24.00	3.50	18.00	34.00	25.00	2.30	20.00	31.00
HR-max Age-related	146.50	13.00	129.00	167.00	146.10	12.85	129.00	169.00

Table 4.2 – Pearson's Correlation coefficients between ZB and FC-HR heart rate variables at Rest.

		ZB vs. FC-HR Heart Rates		
Measure	Time (min: sec)	ZB-M_1 and FC-HR-M_2	Mean Diff. (bpm)	Pearson's Correlations
1st	0:30	71.32 – 71.92	-0.60	0.98**
2nd	1:00	71.62 – 72.00	-0.38	0.98**
3rd	1:30	71.67 – 71.72	-0.05	0.97**
4th	2:00	71.72 – 71.90	-0.18	0.98**
5th	2:30	71.30 – 71.33	-0.03	0.97**
6th	3:00	71.56 – 71.53	0.03	0.99**
7th	3:30	71.62 – 71.78	-0.16	0.98**
8th	4:00	71.50 – 71.95	-0.45	0.98**
9th	4:30	71.73 – 71.87	-0.14	0.98**
10th	5:00	71.35 – 71.67	-0.32	0.97**
11th	5:30	71.63 – 71.87	-0.24	0.98**
12th	6:00	71.67 – 71.65	0.02	0.97**
13th	6:30	71.32 – 71.82	-0.50	0.98**
14th	7:00	71.48 – 71.63	-0.15	0.98**
15th	7:30	71.35 – 71.63	-0.28	0.97**
16th	8:00	71.58 – 71.85	-0.27	0.98**
17th	8:30	71.85 – 71.92	-0.07	0.99**
18th	9:00	71.80 – 72.02	-0.22	0.98**
19th	9:30	71.55 – 72.00	-0.45	0.98**
20th	10:00	71.55 – 71.55	0	0.97**

** *p< 0.05*

Table 4.3 – Pearson's Correlation coefficients between ZB and FC-HR heart rate variables during mCAFT.

Measure	Time (min: sec)	ZB vs. FC-HR Heart Rates: ZB-M_1 and FC-HR-M_2	Mean Diff. (bpm)	Pearson's Correlations
1st	0:30	102.93 – 89.42	13.51	0.89**
2nd	1:00	111.65 – 103.93	7.72	0.95**
3rd	1:30	118.42 – 114.63	3.79	0.98**
4th	2:00	120.87 – 119.45	1.42	0.96**
5th	2:30	122.95 – 121.20	1.75	0.98**
6th	3:00	124.65 – 122.63	2.02	0.98**
7th	3:30	114.88 – 110.14	4.74	0.96**
8th	4:00	124.35 – 118.78	5.57	0.97**
9th	4:30	128.47 – 124.22	4.25	0.98**
10th	5:00	130.33 – 127.59	2.74	0.97**
11th	5:30	132.78 – 129.84	2.94	0.99**
12th	6:00	134.35 – 132.67	1.68	0.99**
13th	6:30	124.51 – 120.87	3.64	0.97**
14th	7:00	131.79 – 127.36	4.43	0.97**
15th	7:30	138.08 – 134.05	4.03	0.93**
16th	8:00	141.87 – 138.56	3.31	0.94**
17th	8:30	143.92 – 141.51	2.41	0.96**
18th	9:00	146.62 – 144.00	2.62	0.95**
19th	9:30	128.63 – 124.88	3.75	0.99**
20th	10:00	139.00 – 135.38	3.62	0.99**
21st	10:30	145.38 – 143.00	2.38	0.99**
22nd	11:00	148.25 – 147.50	0.75	0.99**
23rd	11:30	149.88 – 149.38	0.50	0.99**
24th	12:00	151.50 – 150.88	0.62	0.99**

** *p< 0.05*

Table 4.4 – Pearson's Correlation coefficients between ZB and FC-HR heart rate variables throughout Recovery.

Measure	Time (min: sec)	ZB vs. FC-HR Heart Rates: ZB-M_1 and FC-HR-M_2	Mean Diff. (bpm)	Pearson's Correlations
1st	0:30	132.47 – 131.27	1.20	0.73**
2nd	1:00	113.02 – 113.78	-0.76	0.83**
3rd	1:30	103.28 – 103.82	-0.54	0.84**
4th	2:00	97.30 – 98.73	-1.43	0.83**
5th	2:30	95.55 – 95.00	0.55	0.80**
6th	3:00	91.30 – 90.07	1.23	0.70**
7th	3:30	90.07 – 88.40	1.67	0.80**
8th	4:00	89.95 – 87.37	2.58	0.76**
9th	4:30	86.90 – 85.90	1.00	0.81**
10th	5:00	86.93 – 85.93	1.00	0.81**
11th	5:30	86.27 – 84.92	1.35	0.75**
12th	6:00	86.07 – 83.55	2.52	0.71**
13th	6:30	85.05 – 84.33	0.72	0.82**
14th	7:00	85.82 – 83.73	2.09	0.71**
15th	7:30	84.63 – 83.20	1.43	0.71**
16th	8:00	84.75 – 83.52	1.23	0.75**
17th	8:30	83.97 – 83.05	0.92	0.71**
18th	9:00	83.83 – 82.95	0.88	0.70**
19th	9:30	82.60 – 81.59	1.01	0.70**
20th	10:00	82.77 – 80.72	2.05	0.78**

*** p< 0.05*

Table 4.5 – Pearson's Correlation coefficients between FC-HR and FB-ONE Steps Taken and Energy Expenditure variables during mCAFT.

Variable	FC-HR vs. FB-ONE – Steps Taken and Energy Expenditure		
	FC-HR-M_1 and FB-ONE-M_2	Mean Diff.	Pearson's Correlation
Steps Taken	898.80 – 819.37	79.43	0.98**
Energy Expenditure	131.18 – 92.05	39.13	0.80**

** *p< 0.05*

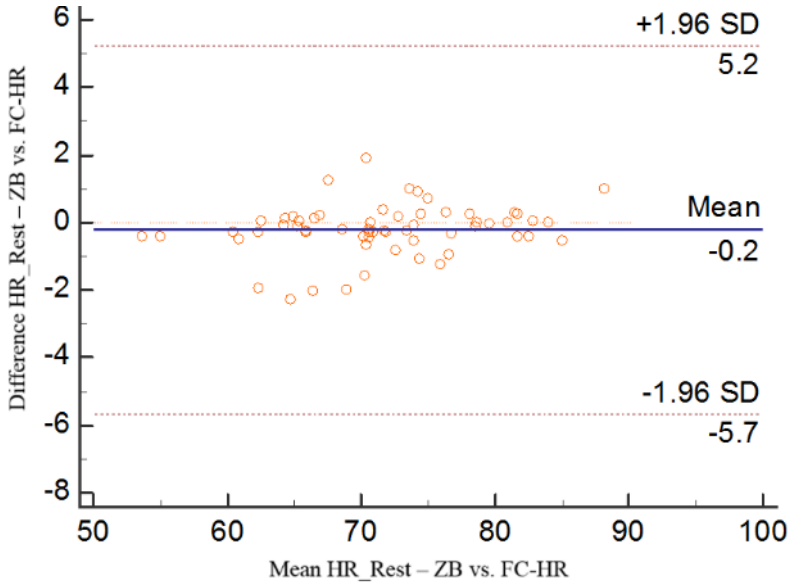
Table – 4.6 One Sample t-test of Mean Difference - FC-HR vs. Adopted Measures.

T-test of Difference	Mean diff.	Std. Error of Mean	95% C I
ZB vs. FC-HR - Rest	-0.20**	0.05	-0.10 – -0.30
ZB vs. FC-HR - mCAFT	4.00**	0.16	3.70 – 4.30
ZB vs. FC-HR - Recovery	1.00**	0.23	0.55 – 1.45
FC-HR vs. F-ONE - ST	79.40**	13.11	53.70 – 105.10
FC-HR vs. F-ONE - EE	39.20**	2.66	34.00 – 44.40

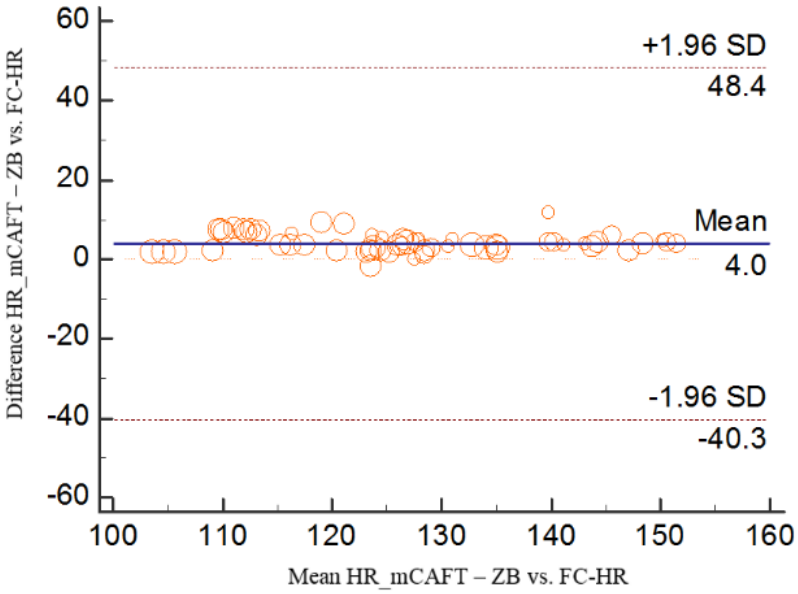
** $p < 0.05$

Figures – 4.1 A-B. Bland Altman plots displaying 95% LoA in pair-wise device comparison.

A

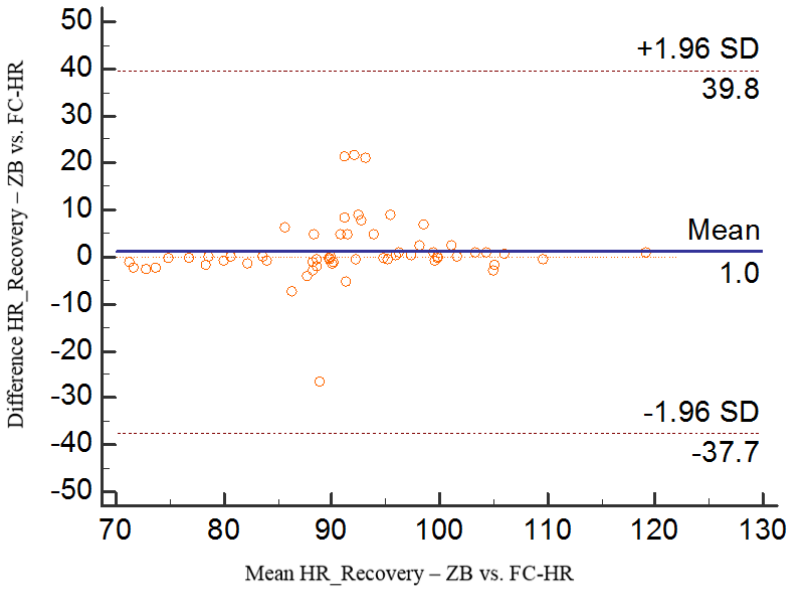


B

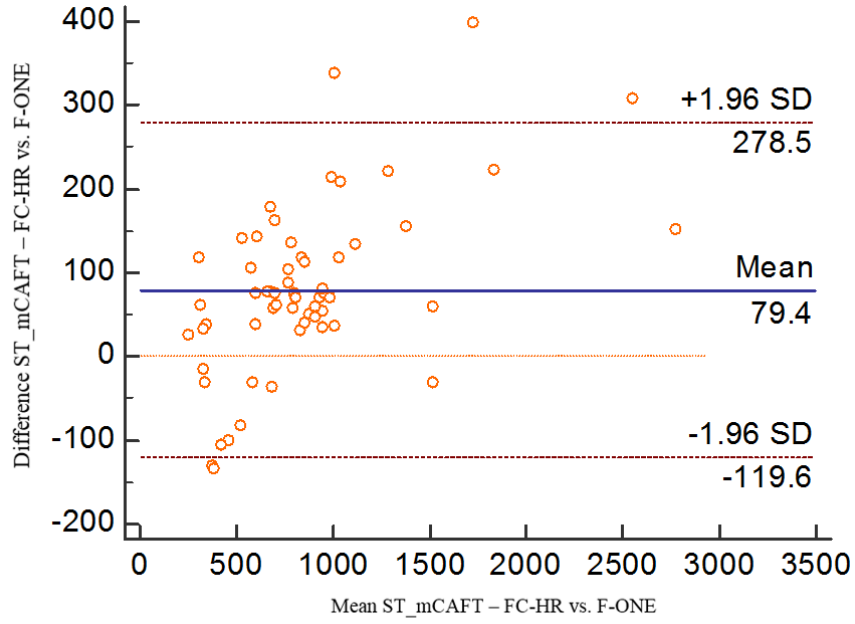


Figures – 4.1 C-D. Bland Altman plots displaying 95% LoA in pair-wise device comparison.

C

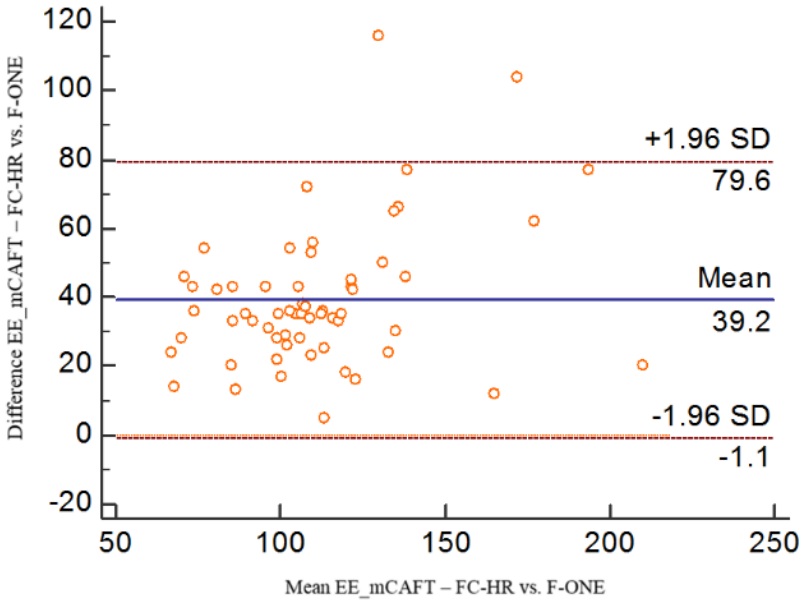


D



Figures – 4.1 E. Bland Altman plots displaying 95% LoA in pair-wise device comparison.

E



CHAPTER 5: DISCUSSION/CONCLUDING REMARKS

This thesis focused on establishing the psychometric and agreement parameters of Zephyr Bioharness (ZB) and Fitbit Charge Heart Rate (FC-HR) wearable physiological monitoring devices. The narrative literature synthesis work conducted on measurement properties of these devices displayed the current existing gaps in the literature regarding the reliability, validity, responsiveness and agreement parameters of ZB and FC-HR devices. Based upon this extensive review, we sampled a large number of participants that included a diverse age range and both genders. We then first addressed the reliability parameters of ZB and FC-HR, and then established the validity and agreement parameters of FC-HR in three distinct phases that also include a sub-maximal fitness test.

This thesis established that excellent reliability measures and strong to very strong correlations can be expected among healthy male and female individuals across various age groups, when using ZB and FC-HR to assess both heart rate and physical activity measures at rest, during the Modified Canadian Aerobic Fitness Test (mCAFT) and throughout recovery. Furthermore, FC-HR and ZB inter-instrument agreement measures of heart rate also demonstrated small average mean biases.

In interpreting our reliability and cross instrument agreement parameters, it is important to consider the context of our study, how measurements were taken, the large variability within our sample, and the nature of the submaximal fitness test used to cause physiologic changes. We divided our studies into three distinct phases to establish the intra-session reliability (consistency) of ZB and FC-HR not just at rest, but also during a sub-maximal activity (mCAFT exercise test), and throughout recovery following cessation of the test. We expected lower ICCs and larger SEM and mean differences during the

transitioning phases of rest to activity and activity to rest, because reliability measures should be assessed during stable conditions and that reliability of a device under unstable conditions would yield lower values. However, the mCAFT administered was a progressively demanding sub-maximal stepping test where the stepping cadence gradually increased as the testing progressed. This steady increase in testing intensity might have contributed to to higher ICCs and lower heart rate mean differences throughout the activity phase of this study. Therefore, our higher and lower reliability measures in more stable and during the times of transition respectively, were both expected and reflects the context of our testing.

While it might be considered a limitation of our reliability study that we actually included phases where we anticipated a lack of stability, this was an inherent to our intention to study the use of devices during a submaximal test. We thought testing in this context was important since submaximal fitness testing can be a valuable aid in physical therapy practice. While we might have all excluded data in periods of the testing where we expected a lack of stability, there is a large inter-subject in these parameters and there was no way to define these a priori. Rather we felt better to evaluate all data, with consideration to the anticipated stability of the activity during which the data was collected. This allowed us to better understand physiological responses during the test itself.

In assessing cross instrument individual agreement parameters between the FC-HR and ZB heart rate measures, small mean differences but wider limits of agreement were observed during the sub-maximal testing and recovery phases. We sampled participants ranging in age from 21 – 68 years and used the mCAFT which requires participants to

achieve their 85% of age related maximum heart rates to terminate the testing. Therefore, the large variability and nature of mCAFT testing could have contributed to these wider individual limits of agreement.

There are many potential applications for wearable sensors in physical therapy practice. Physical therapists prescribe exercise or activity in different contexts including health promotion where the goal is to increase overall physical activity and exercise, to therapeutic exercise where the goal is remediation of impairments. Further, since physical therapists are increasingly confronted with patient populations with comorbid health problems that require exercise, physiological monitoring for safety reasons is also an important application. Since physical therapists frequently engage in assessment of movement and prescription of activity and exercise, these devices have substantial implications for physical therapy practice. It might be anticipated that better monitoring would allow physiotherapists to more accurately assess physiological status and response to exercise in a way that would allow for better diagnosis and progression of exercise interventions. There is evidence suggesting that monitoring improve the adherence. Adherence to exercise or physical therapy home programs is a substantial issue in optimizing the effectiveness of physical therapy. While there are many potential applications for these devices in physical therapy and there is emerging evidence that supports the use of these devices, much research will be needed before the full range of potential applications can be fully elucidated in the literature.

The strengths of this thesis included that fact that we established the psychometric and agreement properties of multiple devices at stable and progressive sub-maximal

conditions, and sampled male and female participants with a diverse age range. Despite this, in our reliability study, we did not address ahead of time that lower reliability measures would be expected during the transitioning phases. However, it would be extremely difficult to establish the exact steady-state within the sub-maximal and recovery phases. We also did not assess the performance of FC-HR measures with gold standard (calibration), but rather used previously established reliable and valid devices with good agreements parameters with gold standard measures.

We studied the reliability, validity and agreement properties of ZB and FC-HR devices during human performance in healthy participants, however, did not establish these parameters in any specific patient population. Future studies assessing both the psychometric and agreement parameters of the two devices in different patient populations or in various occupational contexts are warranted. This is particularly important in physical therapy practice where it is more common to be managing patient populations, then primary prevention in clients without pathology.

An important distinction when conducting studies on wearable technologies is the difference between the reliability and validity of the devices and their component sensors versus the reliability and validity of the data from patient protocols. While there is some interdependence between these two characteristics, one does not necessarily guarantee the other. A sensor may provide reliable data on a patient test protocol without actually reflecting the true physiological measure that it is intended to do due to technical limitations in the sensors. Conversely, a sensor may have the capacity to provide reliable and valid data, but if the protocol by which the device is used for testing is not appropriate or

consistent, these measurement properties will not be achieved in the patient testing protocol. This thesis focused on the reliability and validity of wearable technologies in the patient testing protocol. Since the devices were not calibrated against a gold standard to determine their criterion validity, we are unable to confirm the ability to accurately measure physiological property, or to compare different devices in this regard.

Our study findings have significant implications in the field of continuous heart rate monitoring. By establishing the reliability and validity measures of ZB and FC-HR monitoring devices, researchers could obtain more consistent and accurate real-time field-based wireless measures of heart rate using small, portable and less costly devices, throughout both stable and sub-maximal conditions. Furthermore, assessing the agreement parameters between FC-HR and ZB heart rate measures, provided information regarding a possible the interchangeable use.

In conclusion, this thesis adds to the existing pool of literature regarding the psychometric and agreement parameters of wearable devices in motoring of physiologic measures.

Appendix – A: Sample Size Calculation for Relative Reliability Hypothesis Testing:

$$n = \frac{0.5\ k\ (Z_{\alpha} + Z_{\beta})^2}{\delta^2\ (k - 1)} + 2$$

- Number of occasions, k=2
- Z_{α} is the tabled Z-value associated with the α value of interest. Z-value (1-tailed) for of 0.05 is equal to 1.645
- Z_{β} is the Z-value associated with a Type II error. The Z-value (1-tailed) for β of 0.20 is equal to 0.842.
- δ is the difference between null hypothesis Z transformed R-value and the expected Z transformed R-value.

  $\delta = Z_{Rexpected} - Z_{Rnull}$

- $Z_{R\text{-}expected}$ is the Z value associated with the reliability you hope to obtain in your study,

$$Z_{Rexpected} = .5\,\text{natural log}\frac{1 + (k - 1)\,R_{expected}}{1 - R_{expected}}$$

- $Z_{R\text{-}lower\ limit}$ is the lower confidence limit for the desired confidence interval width.

$$Z_{Rlowerlimit} = .5\,\text{natural log}\frac{1 + (k - 1)\,R_{lowerlimit}}{1 - R_{lowerlimit}}$$

Calculations:

Step 1:

$$Z_{Rexpected} = .5 \, \text{natural log} \frac{1 + (2 - 1)\, 0.90}{1 - 0.90}$$

$$Z_{Rexpected} = 1.47$$

Step 2:

$R_{lowerlimit} = 0.90 - 0.10$
$R_{lowerlimit} = 0.80$

Step 3:

$$Z_{Rnull} = .5 \, \text{natural log} \frac{1 + (2 - 1)\, 0.80}{1 - 0.80}$$

$$Z_{Rnull} = 1.09$$

Step 4:

$\delta = 1.47 - 1.09 = 0.38$

$\delta^2 = 0.38^2 = 0.14$

Step 5:

$$n = \frac{0.5\ k\ (Z_{\alpha} + Z_{\beta})^2}{\delta^2\ (k-1)} + 2$$

$$n = \frac{0.5\ (2)\ (1.645 + 0.842)^2}{(0.14)\,(2-1)} + 2$$

$$n = 46 \text{ patients}$$

Step 6:

Estimating a 20% rate of expected drop out:

$$n = \frac{46}{0.80} = 57.5$$

A sample size of **60** participants will be required.

Appendix – B: Participant Information Sheet and Consent Form

PARTICIPANT INFORMATION SHEET

Title of Study:	**The use of Zephyr Bioharness to determine the physiological changes, and to establish the psychometric properties of Fitbit Charge Heart Rate, in healthy participants during the modified Canadian Aerobic Fitness Test.**
Local Principal Investigator:	**Dr. Joy MacDermid, PhD, McMaster University.**
Principal Investigator:	**Mr. Goris Nazari, PT, MSc (C), School of Rehabilitation Science McMaster University.**

You are being invited to participate in a research study led by Mr. Goris Nazari and Drs. Joy MacDermid & Kathryn Sinden.

In order to decide whether or not you want to be a part of this research study, you should understand what is involved and the potential risks and benefits. This form gives detailed information about the research study. Please take your time to make your decision. Feel free to discuss it with your friends, family, and healthcare providers.

WHY IS THIS RESEARCH BEING DONE?

It is important to obtain accurate heart rates and breathing rates of health individuals as well as patients at home or in a hospital setting. The traditional manual pulse palpation for heart rate and directly observing the number of breaths per minute for breathing rate recordings have been considered to be not suitable.

Similarly monitoring the number of steps taken and the amount of energy expended are both beneficial in terms of measuring ones physical activity levels since physical inactivity and obesity both impose a major burden on the Canadian health care system.

WHAT IS THE PURPOSE OF THE STUDY?

1-To record the participants' heart rate and breathing rate during a stair climbing activity using a device called Zephyr Bioharness. The Zephyr BioHarness is a U.S. FDA-approved wireless monitoring device that records a one's heart rate and breathing rate. The Zephyr Bioharness consists of an adjustable chest belt with grey colored sensors and an electronic BioModule that snaps on to the belt. The device is fitted comfortably on the chest at the lower sternum for both men and women and weighs 85grams.

2-To record participants' the heart rate, number of steps taken and total amount of energy expended using the Fitbit Charge Heart Rate and Fitbit-One devices during a stair climbing activity. Fitbit Charge Heart Rate (Fitbit Inc., San Francisco, CA) is a wristband made of a flexible material which would fit wrists 13.7 to 22.1 cm. It includes a heart rate monitor and an accelerometer to provide number of step counts and the total amount of energy expended. The device is sweat, rain and splash proof. Fitbit One (Fitbit Inc., San Francisco, CA) is a small and lightweight advanced accelerometry-based device which can be worn in the front pocket of shorts and it tracks physical activity. Its physical activity recording features include; number of steps taken and energy expenditure.

WHAT WILL MY RESPONSIBILITIES BE IF I TAKE PART IN THE STUDY?

If you volunteer to participate in study, you will be asked to schedule two sessions and come in to the MacHand Lab - 310A, Institute of Applied Health Sciences Building, 72 hours (three days) apart.

<u>During your first visit</u>

A-You will be asked to complete the followings:

1) Sign a Consent Form,
2) Physical Activity Readiness Questionnaire,

B-Asked to go through the Screening procedures for recording of:

1) Resting Heart Rate,
2) Resting Blood Pressure,
3) Height (m),
4) Weight (kg),

C-Fitting of Zephyr and Fitbit devices:

1) Zephyr device: The device will be fitted comfortably on the chest at the lower sternum (xiphoid process) for both men and women and weighs 85-grams.
2) Fitbit Charge Heart Rate: This wristband will be fitted on each participants' right wrist.
3) Fitbit-One: This device will be worn in the front pocket of each participants' shorts.

D- You will be familiarized with the Modified Canadian Aerobic Fitness Test.

E-The Modified Canadian Aerobic Fitness Test will be administered.

During you second visit

A-Asked to go through the Screening procedures for recording of:

1) Resting Heart Rate,
2) Resting Blood Pressure,

B-Fitting of Zephyr and Fitbit devices:

1) Zephyr device: The device will be fitted comfortably on the chest at the lower sternum (xiphoid process) for both men and women and weighs 85-grams.
2) Fitbit Charge Heart Rate: This wristband will be fitted on each participants' right wrist.
3) Fitbit-One: This device will be worn in the front pocket of each participants' shorts.

C-The Modified Canadian Aerobic Fitness Test will be administered.

WHAT ARE THE POSSIBLE RISKS AND DISCOMFORTS?

The Zephyr Bioharness chest strap and Fitbit Charge Heart Rate wrist band may feel uncomfortable while performing the modified Canadian Aerobic Fitness Test. However, the principal investigator will ask for each participant's feedback and adjust the chest strap and the wristband accordingly.

The testing procedures will be very similar to what you do on daily basis which is simple stair climbing, however it will be more challenging and you will experience a normal increase of heart rate and breathing rate while performing the test. This will be considered normal. The total testing duration will be for less than twelve minutes with approximately three one-minute rest periods. Once the test is completed, you will be asked to walk around the room for two minutes and then asked to be seated and your resting heart rate will be recorded at 2:30, 3:30 and 4:30 minutes, post- step test termination to ensure that the your heart rates drops below 100 beats per minute before you leave the testing site.
Therefore, it is very unlikely that you will have any difficulties or get hurt when doing this test. If a problem occurs during the study protocol, standard emergency response procedures will be invoked and implemented.

HOW MANY PEOPLE WILL BE IN THE STUDY?

There will be a total of 60 participants.

WHAT ARE THE POSSIBLE BENEFITS FOR ME AND/OR FOR SOCIETY?

We cannot promise any personal benefits to you from your participation in this study. We hope to learn more about wireless monitoring of heart rate, breathing and physical activity levels. However the results of this study may benefit society and the scientific community by developing a reliable and valid tool that may help monitor heart rate, breathing rate and physical activity levels in the hospital, athletic and recreational settings.

WHAT INFORMATION WILL BE KEPT PRIVATE?

You are participating in this study confidentially. I will not use your name or any information that would allow you to be identified. No one but me and the research coordinator will know whether you participated unless you choose to tell them. Every effort will be made to protect your confidentiality and privacy. The information you provide will be kept in a locked cabinet where only I and the research coordinator will have access to it. Once the study is complete, an archive of the data, without identifying information, will be deposited.

For the purposes of ensuring the proper monitoring of the research study, it is possible that a member of the Hamilton Integrated Research Ethics Board may consult your research data. However, no records which identify you by name or initials will be allowed to leave the University. By signing this consent form, you or your legally acceptable representative authorize such access.

If the results of the study are published, your name will not be used and no information that discloses your identity will be released or published without your specific consent to the disclosure.

CAN PARTICIPATION IN THE STUDY END EARLY?

If you volunteer to be in this study, you may withdraw at any time. You have the option of removing your data from the study. You may also refuse to answer any questions you don't want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

WILL I BE PAID TO PARTICIPATE IN THE STUDY?

You will only be reimbursed for your parking expenses when you come for your

lab visits.

WILL THERE BE ANY COSTS?

There are no costs associated with this study.

WHAT HAPPENS IF I HAVE A RESEARCH-RELATED INJURY?

If you are injured as a direct result of taking part in this study, all necessary medical treatment will be made available to you at no cost. Financial compensation for such things as lost wages, disability or discomfort due to this type of injury is not routinely available.

However, if you sign this consent form it does not mean that you waive any legal rights you may have under the law, nor does it mean that you are releasing the investigator(s), institution(s) and/or sponsor(s) from their legal and professional responsibilities.

IF I HAVE ANY QUESTIONS OR PROBLEMS, WHOM CAN I CALL?

If you have any questions about the research now or later, please contact:

Principal Investigator: Mr. Goris Nazari

Cell Phone: 905-923-0748 or E-mail: nazarigs@mcmaster.ca

Local Principal Investigator: Dr. Joy MacDermid

Phone: 905-525-9140 ext. 22524 or E-mail: macderj@mcmaster.ca

This study has been reviewed by the Hamilton Integrated Research Ethics Board (HIREB). The HIREB is responsible for ensuring that participants are informed of the risks associated with the research, and that participants are free to decide if participation is right for them. If you have any questions about your rights as a research participant, please call the Office of the Chair, Hamilton Integrated Research Ethics Board at 905.521.2100 x 42013.

CONSENT STATEMENT

Participant:

I have read the preceding information thoroughly. I have had an opportunity to ask questions and all of my questions have been answered to my satisfaction. I agree to participate in this study. I may withdraw from the study at any time.I understand that I will receive a signed copy of this form.

Name	Signature	Date

Person obtaining consent:

I have discussed this study in detail with the participant. I believe the participant understands what is involved in this study.

Name, Role in Study	Signature	Date

Contact for Future Research Project Opportunities

I agree to be contacted about future research and
I understand that I can always decline the request. Yes No

Please contact me at: _____

_____ _____ _____

Name of Participant (Printed) Signature Date

Appendix – C: Physical Activity Readiness Questionnaire (PAR-Q)

Physical Activity Readiness Questionnaire - PAR-Q (revised 2002)

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	NO		
☐	☐	1.	**Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?**
☐	☐	2.	**Do you feel pain in your chest when you do physical activity?**
☐	☐	3.	**In the past month, have you had chest pain when you were not doing physical activity?**
☐	☐	4.	**Do you lose your balance because of dizziness or do you ever lose consciousness?**
☐	☐	5.	**Do you have a bone or joint problem (for example, back, knee or hip) that could be made worse by a change in your physical activity?**
☐	☐	6.	**Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?**
☐	☐	7.	**Do you know of any other reason why you should not do physical activity?**

If you answered

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.

CSEP | SCPE

Appendix – D: Critical Appraisal of Study Design for Psychometric Articles: Evaluation Form

		Evaluation Criteria	Score		
Study question			**2**	**1**	**0**
1		Was the relevant background research cited to define what is currently known about the psychometric properties of the measures under study and the need or potential contributions of the current research question?			
Study design					
2		Were appropriate inclusion/exclusion criteria defined?			
3		Were specific psychometric hypotheses identified?			
4		Was an appropriate scope of psychometric properties considered?			
5		Was an appropriate sample size used?			
6		Was appropriate retention/follow-up obtained? (Studies involving retesting or follow-up only)			
Measurements					
7		Documentation: Were specific descriptions provided or referenced that explain the measures and their correct application/interpretation (to a standard that would allow replication)?			
8		Standardized methods: Were administration and application of measurement techniques within the study standardized, and did they consider potential sources of error/misinterpretation?			
Analyses					
9		Were analyses conducted for each specific hypothesis or purpose?			
10		Were appropriate statistical tests conducted to obtain point estimates of the psychometric property?			
11		Were appropriate ancillary analyses done to describe properties beyond the point estimates (confidence intervals [CI], benchmark comparisons, standard error of measurement [SEM], minimally important difference [MID])?			
Recommendations					
12		Were the conclusions/clinical recommendations supported by the study objectives, analysis, and results?			
Subtotal (of columns 1 & 2)					
Total score % (sum of subtotals/24 × 100), or, if for a specific paper an item is deemed inappropriate, then sum items, divide by 2 times the number of items, and multiply by 100 to get the percentage score.					

Appendix – E: Preliminary Participation Instruction Form

Preliminary Participant Instructions

Title of Study:	**The use of Zephyr Bioharness to determine the physiological changes, and to establish the psychometric properties of Fitbit Charge Heart Rate, in healthy participants during the modified Canadian Aerobic Fitness Test.**
Local Principal Investigator:	**Dr. Joy MacDermid, PT, PhD, School of Rehabilitation Science McMaster University.**
Principal Investigator:	**Mr. Goris Nazari, PT, MSc (C), School of Rehabilitation Science McMaster University.**

Please adhere to the following instructions prior to your sessions:

- Dress requirements: Shorts and short-sleeved or sleeveless shirt should be worn. Running shoes are the recommended footwear. The Zephyr Bioharness chest strap will be placed next to the skin on the chest at the lower sternum (Xiphoid process) under sport bras or athletic tops. The FitBit Charge Heart Rate will be worn on your right wrist and Fitbit-One will be clipped on to the front pocket of your shorts.
- Food and Beverages: Do not eat for at least two hours prior to your appraisal. Also refrain from drinking caffeine beverages for two hours and alcoholic drinks for six hours prior to the appraisal.
- Smoking: Do not smoke during the two hours prior to the appraisal.
- Physical Activity: Strenuous physical activity should be avoided for six hours prior to the appraisal.