# EVIDENCE ANALYSIS PROJECT:

# VALIDITY AND UTILITY OF ACTIVITY MONITORS TO IMPROVE HEALTH

## OUTCOMES

Alana Degeneffe, RDN, CD

Submitted in partial fulfillment of the requirements for the degree

Master of Science in Dietetics

Mount Mary University

December 2018

Approved by

Megan Baumler, PhD, RD Professor, Graduate Program in Dietetics

Tara LaRowe, PhD, RDN Professor, Graduate Program in Dietetics

# EVIDENCE ANALYSIS PROJECT: VALIDITY AND UTILITY OF ACTIVITY MONITORS TO IMPROVE HEALTH OUTCOMES

# Alana Degeneffe, RDN, CD

# December 2018

# ABSTRACT

BACKGROUND: Worldwide obesity rates have nearly tripled since 1975 and are projected to increase further. The prevention and treatment of obesity has focused on pharmacological, educational, and behavioral interventions with limited success overall. A recent trend and possible technological intervention, activity monitors are wearable accelerometer-based devices aimed for the consumer market. Activity monitoring technology offers promise for improving adherence and weight loss outcomes.

OBJECTIVE: To critically analyze and appraise current evidence on the validity and utility of activity monitors to improve health outcomes in adults.

DESIGN: The Academy of Nutrition and Dietetics' (AND) Evidence Analysis Library (EAL) project.

METHODS: The five steps included in the AND's Evidence Analysis process are 1) Formulate the Evidence Analysis Question, 2) Gather and Classify the Evidence, 3) Critically Appraise Each Article, 4) Summarize the Evidence, and 5) Write and Grade the Conclusion Statement.

RESULTS: Using the PubMed database, a total of 18 studies were identified relating to activity monitor validity and utility in adults. Six studies were excluded because they assessed participants with more than one health condition, the predictability of body weight changes, or activity monitors that were not consumer-grade devices. Twelve studies were included for further analysis. Of the 12 studies, seven were validity studies and five were randomized controlled trials, non-randomized crossover trials, and a cross sectional study exploring activity monitor utility. Not all studies analyzed activity monitor validity or utility in improving health outcomes within the same parameters. Four validity studies were conducted under controlled, research environments utilizing gold standard comparisons, while three validity studies were conducted in free-living environments as activity monitors are intended for, utilizing common field-based devices as comparisons. Utility studies did not evaluate activity monitors as independent interventions. Instead, most utility studies incorporated activity monitors as self-monitoring tools integrated into behavioral interventions with positive outcomes.

CONCLUSION: Current, consumer-grade activity monitors exhibit moderate validity on average, tend to estimate step counts accurately, underestimate heart rate and energy expenditure, overestimate time asleep, and are more accurate at rest than during activity. Adults who utilize current, consumer-grade activity monitors as combined interventions may experience

a clinically meaningful increase in steps, physical activity, and weight loss. This conclusion was graded *II*, *Fair*.

# ACKNOWLEDGMENTS

I would like to thank my parents, grandparents, and fiancé for instilling a fierce work ethic in me, believing I can achieve anything I set my mind to, and encouraging me through my entire academic journey. I would not be where I am today without your love and support.

Thank you to my colleagues and clients for inspiring me to pursue this research topic. Last but certainly not least, thank you to Megan Baumler, my professor and advisor for sticking with me through this process. I truly appreciate all the time and effort you dedicated to guiding and improving my Evidence Analysis project.

Chapte	r	Page
I.	Introduction	5
II.	Literature Review	10
III	Methodology	
IV	Results	
V.	Conclusion	44
Biblio	graphy	
Appen	dices	
A.	Quality Criteria Checklists and Evidence Analysis Library Worksheets	53
B.	Overview Table	
C.	Comparison Report	

# TABLE OF CONTENTS

## CHAPTER I: INTRODUCTION

Overweight and obesity are defined as abnormal, excessive fat accumulation that may impair health. In 2016, nearly two billion adults (39 percent of the world's population) were overweight. Of these, over 650 million adults (13 percent of the world's population) were obese. The prevalence has also risen among children. In 2016, over 18 percent of the world's population of children were overweight or obese. Since 1975, obesity rates have nearly tripled, and are projected to increase further (World Health Organization, 2018).

Personal behaviors play a dominant role in preventing and treating obesity (Hruby & Hu, 2015), as the fundamental cause of obesity is an energy imbalance between calories consumed and calories expended. Lifestyle adaptations can result in energy imbalances that promote obesity, such as consuming calorically-dense diets without adequate physical activity. Physical activity is defined as any body movement produced by the skeletal muscles that requires energy expenditure. Therefore, physical activity is an essential component of weight management. However, an estimated 31 percent of adults worldwide do not meet the recommended levels. Physical inactivity contributes to obesity, chronic illnesses, and other healthcare concerns (World Health Organization, 2018), while obesity has drastic impacts on morbidity, mortality, and economic burden (Hruby & Hu, 2015).

A recent trend and possible technological intervention, activity monitors are wearable accelerometer-based devices aimed for the consumer market. Most activity monitors have displays for immediate health feedback, including step counts, elevation or stairs climbed, distance traveled, heart rate, calories burned, active time, and time asleep (Ferguson, Rowlands, Olds, & Maher, 2015). Additionally, many activity monitors have associated websites, smartphone applications, and smart scales that synchronize with the wearable devices. Users can

track their workouts, water consumption, caloric intake, and weight fluctuations through the synchronizing options (Ross & Wing, 2016).

Technology-based tools create awareness by assisting individuals in tracking their weight-related behaviors (Ross & Wing, 2016). Activity monitoring technology shows potential for facilitating self-motivation, self-monitoring, self-efficacy, and positive behavior change, all of which are essential qualities for successful behavioral weight management treatments (Ferguson et al., 2015). Overall, activity monitoring technology offers promise for improving adherence and weight loss outcomes (Ross & Wing, 2016). However, the usefulness of activity monitors depends highly on their accuracy.

# Rationale

The prevention and treatment of obesity has focused on pharmacological, educational, and behavioral interventions, with limited success overall (Townshend & Lake, 2017). However, the implementation of physical activity programs with self-motivating, self-monitoring mechanisms may be efficacious in suppressing the global physical inactivity and obesity epidemics. Furthermore, efforts to increase physical activity and control obesity will result in tremendous economic savings and remarkable health benefits (Lewis, Lyons, Jarvis, & Baillargeon, 2015). The purpose of this Evidence Analysis project is to critically analyze and appraise current evidence on the validity and utility of activity monitors to improve health outcomes in adults. A systematic review of existing literature has been conducted, and conclusion statements have been determined based on the findings of the review.

## **Potential Significance**

Results from this Evidence Analysis project will indicate whether activity monitors are useful additions to healthy lifestyle interventions. The findings may also contribute to future

research and product development. With progress, activity monitor functions should reach reference device standards, and may have practical clinical applications for all populations. Dietitians and other healthcare professionals could provide evidence-based recommendations regarding appropriate activity monitor selection. With automatic data transfer of validated activity monitors, healthcare professionals could accurately track clients' physical activity habits, heart rate, energy expenditure, sleep patterns, food and beverage consumption, and weight fluctuations. Direct access to such data would save an immense amount of time, provide a holistic picture of health for each client, and enable healthcare professionals to personalize care more than ever before. Future technologies may be able to provide an early warning of disease, aid in diagnosis and treatment, and contribute to a deeper understanding of human health (Savage, 2017). Hopefully, the individualized approach to care by means of utilizing activity monitor data will result in better health outcomes for all populations, ultimately leading to a decreased prevalence of overweight and obesity, chronic illnesses, mortality, and healthcare costs.

## **Objectives**

- A. Identify the validity of current, consumer-grade activity monitor functions
- B. Identify the health outcome improvements when current, consumer-grade activity monitors are utilized
- C. Compare 10 current, best-selling, commercially-available activity monitor product details in a supplemental comparison report

## **Research Questions**

1. How valid are the functions of current, consumer-grade activity monitors compared to research-grade devices?

2. Does physical activity improve and weight loss occur as a result when adults utilize current, consumer-grade activity monitors?

## **Research Hypotheses**

- 1. Current, consumer-grade activity monitor functions will be within 25 percent error range of research-grade devices
- 2. Adults who utilize current, consumer-grade activity monitors will experience an increase in physical activity and weight loss as a result

## Limitations

Data has been collected exclusively from published research. Due to the nature of an Evidence Analysis project, two limitations are the reliance on, and lack of pertinent research articles. In regard to objective C of this Evidence Analysis project, financial constraint was an additional limitation, restricting data collection to manufacturer websites only.

#### **Delimitations**

Data has been collected from the most current studies due to the rapidly evolving activity monitor market. Validity data was limited further to the activity monitors and specific functions tested in each study, even though updated activity monitor models may currently be on the market. Likewise, utility data must take the results, strengths, and limitations of activity monitor validity data into consideration.

## Assumptions

This Evidence Analysis project assumes the accuracy and honesty of all published research.

#### **Definition of Terms**

- Accelerometry: the electromechanical measurement of acceleration and deceleration in a part of, or the entire body during the performance of a task. A common, noninvasive procedure used to capture physical activity intensity using a wearable device called an accelerometer.
- Actigraphy: a method of monitoring body movements over time to determine periods of rest versus activity. A common, noninvasive procedure used to detect sleep disorders by using a wearable device called an actigraph.
- Electrocardiography: a noninvasive procedure used to record electrical changes in the heart. The record, which is called an electrocardiogram, shows the series of waves that relate to the electrical impulses that occur during each beat of the heart.
- **Indirect calorimetry**: a noninvasive procedure used to quantify energy expenditure. An individual's heat production is determined by measuring oxygen uptake and carbon dioxide output over a given period of time.
- **Polysomnography**: a noninvasive procedure used to study sleep and diagnose sleep disorders. The record, which is called a polysomnogram, shows brain waves, blood oxygen levels, heart and breathing rates, eye and leg movements, and sleep stages.
- Utility: the state of making practical and effective use of something
- Validity: the extent to which a variable or measure captures the concept it is intended to reflect

## CHAPTER II: LITERATURE REVIEW

## Introduction

Overweight and obesity are defined as abnormal, excessive fat accumulation that may impair health (World Health Organization, 2018). The current, most widely used criteria for classifying overweight and obesity is body mass index (BMI), which is calculated by dividing body weight in kilograms by height in meters squared (Hruby & Hu, 2015). A BMI greater than or equal to 25 is considered overweight. A BMI greater than or equal to 30 is considered obese. In 2016, nearly two billion adults (39 percent of the world's population) were overweight. Of these, over 650 million adults (13 percent of the world's population) were obese. The prevalence has also risen among children. In 2016, over 18 percent of the world's population of children were overweight or obese. Obesity rates have nearly tripled since 1975 and are projected to increase further (World Health Organization, 2018).

The fundamental cause of overweight and obesity is an energy imbalance between calories consumed and calories expended (World Health Organization, 2018). This energy imbalance is partially a result of environmental changes beyond the control of any individual. Environmental changes that promote obesity, known as obesogenic environments, have been fueled by industrialization, automation, transportation, urbanization, economic growth, sedentary lifestyles, and consuming highly processed, calorically-dense diets. Risk factors of obesity unrelated to environmental changes include lower socioeconomic status, limited education, and hereditary factors such as genetics, family history, racial, and ethnic differences (Hruby & Hu, 2015). Likewise, obesity is a major risk factor for a number of noncommunicable diseases, namely musculoskeletal disorders, diabetes, cardiovascular disease, and certain cancers. As BMI increases, the risk for theses chronic diseases also increases (World Health Organization, 2018). Overall, obesity has drastic impacts on morbidity, mortality, healthcare costs, and economic burden. However, obesity risk factors are modifiable. In response to these conditions, personal behaviors play a dominant role in preventing and treating obesity (Hruby & Hu, 2015).

Physical activity is defined as any body movement produced by the skeletal muscles that requires energy expenditure (World Health Organization, 2018). Therefore, physical activity is an essential component of weight management. According to the 2008 Physical Activity Guidelines for Americans, adults 18 to 64 years of age should do at least 150 minutes of moderate-intensity, or 75 minutes of vigorous-intensity aerobic physical activity per week (Office of Disease Prevention and Health Promotion, 2018). However, it is estimated that 31 percent of adults worldwide do not meet these recommended levels. Physical inactivity can be attributed to an increase in the use of transportation, sedentary behavior during occupational and domestic activities, and insufficient participation in physical activity during leisure time. Strong evidence demonstrates that adults who meet the recommended levels of physical activity have lower rates of fractures, depression, diabetes, cardiovascular disease, cancer, and mortality compared to those who do not. Adults who are more active also exhibit a higher level of physical fitness, a lower BMI, and are more likely to achieve weight maintenance (World Health Organization, 2018). Thus, efforts to increase physical activity and control obesity will result in tremendous economic savings and remarkable health benefits (Lewis, Lyons, Jarvis, & Baillargeon, 2015). The implementation of physical activity programs with self-motivating, selfmonitoring mechanisms may be efficacious in suppressing the global physical inactivity and obesity epidemics.

A recent trend and possible technologic intervention, activity monitors are wearable accelerometer-based devices aimed for the consumer market. Most activity monitors have

displays for immediate health feedback, including step counts, elevation or stairs climbed, distance traveled, heart rate, calories burned, active time, and time asleep (Ferguson, Rowlands, Olds, & Maher, 2015). Additionally, many activity monitors have associated websites, smartphone applications, and smart scales that synchronize with the wearable devices. Users can track their workouts, water consumption, caloric intake, and weight fluctuations through the synchronizing options (Ross & Wing, 2016).

Activity monitoring technology also shows potential for facilitating self-motivation, selfmonitoring, self-efficacy, and positive behavior change, all of which are essential qualities for successful behavioral weight management treatments (Ferguson et al., 2015). Technology-based tools create awareness by assisting individuals in tracking their weight-related behaviors. Activity monitoring technology offers promise for improving adherence and weight loss outcomes (Ross & Wing, 2016). However, the usefulness of activity monitors depends highly on their accuracy. The purpose of this literature review is to critically analyze current evidence on the validity and utility of activity monitors to improve health outcomes in adults.

## Background

The prevention and treatment of obesity has focused on pharmacological, educational, and behavioral interventions, with limited success overall (Townshend & Lake, 2017).

# Pharmacotherapy

The history of weight loss drugs has seen the rise and fall of numerous medications that proved highly effective, but ultimately dangerous. The purpose of pharmacotherapy is not to search for a magic pill, but rather to apply a safe and effective drug regimen, in combination with improved diet and exercise to achieve a sustainable reduction in body weight. As of 2013, only three drugs were approved by the United States Food and Drug Administration (FDA) as adjunctive therapy for chronic weight management. The FDA-approved anti-obesity medications include orlistat (brands include Alli and Xenical), lorcaserin (trade name Belviq), and phentermine/topiramate extended-release (trade name Qsymia) (Kim, Lin, Blomain, & Waldman, 2014).

Orlistat is the only FDA-approved anti-obesity medication that is available without a prescription. Orlistat acts by binding and inhibiting pancreatic and gastrointestinal lipases from breaking down dietary triglycerides into free fatty acids, which can be absorbed via fatty acid transporters expressed by the intestinal epithelial cells. Thus, orlistat decreases systemic fat absorption and reduces caloric intake. However, major adverse effects of orlistat include steatorrhea and the risk of fat-soluble vitamin deficiencies. Hepatotoxicity, nephrotoxicity, pancreatitis, and kidney stones are additional safety concerns (Kim, Lin, Blomain, & Waldman, 2014).

Serotonin is a neurotransmitter that mediates several processes in the central nervous system. One of the physiological processes that serotonin regulates is postprandial satiety through hypothalamic serotonin receptors. Lorcaserin is a selective hypothalamic serotonin receptor agonist. Therefore, lorcaserin decreases appetite and food intake, prevents weight gain, and assists in weight loss. Although, adverse effects of lorcaserin include headache, dizziness, fatigue, dry mouth, nausea, and constipation. Psychiatric disorders, cardiovascular events, and carcinogenesis are additional safety concerns (Kim, Lin, Blomain, & Waldman, 2014).

Given the complex, multifactorial etiology of obesity, it is unlikely that one weight loss drug will be sufficient to reverse the condition. Unsurprisingly, combination therapies such as phentermine/topiramate extended-release have been evaluated and show a greater potential in the treatment of obesity. Phentermine's mechanism of action is reliant on modulation of

catecholamines in the satiety centers of the hypothalamus, thus decreasing appetite. Topiramate is an alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate/kainate receptor antagonist. Modification of voltage-gated sodium and calcium channels, as well as induction of gammaaminobutyric acid receptor-mediated inhibitory currents may also contribute to topiramate's weight loss effects by decreasing food intake and efficiency of nutrient utilization, and increasing energy expenditure. However, various adverse effects of phentermine/topiramate extendedrelease have been noted, such as headache, dizziness, blurry vision, paresthesia, insomnia, anxiety, depression, dry mouth, upper respiratory tract infection, nasopharyngitis, and constipation. Psychiatric disorders, cardiovascular events, and teratogenesis are additional safety concerns. The modest efficacy, undesirable adverse effects, serious health risks, questionable safety, and potential of weight regain combine to highlight the major limitations of weight loss drugs in the treatment of obesity (Kim, Lin, Blomain, & Waldman, 2014).

## Medical Nutrition Therapy

The AND describes medical nutrition therapy (MNT) as an evidence-based, in-depth application of the Nutrition Care Process (NCP). The NCP entails an individualized nutrition assessment, determination of the nutrition diagnosis, determination and application of the nutrition intervention appropriate for the individual or group, routine monitoring, and evaluation to manage the disease, condition, or injury. MNT employs all domains of nutrition intervention, including food and nutrient delivery, nutrition education, nutrition counseling, and coordination of nutrition care. MNT services are provided by registered dietitians for individuals and groups utilizing meal plans, medically prescribed diets, specialized oral feedings, tube feedings, intravenous solutions, and the analysis of potential food and drug interactions. The provision of MNT typically results in the prevention, delay, or management of diseases, conditions, or injuries (Academy of Nutrition and Dietetics, 2017).

The AND defines nutrition education as the reinforcement of basic or essential nutritionrelated knowledge (Academy of Nutrition and Dietetics, 2006). In regard to weight loss and maintenance, nutrition education may cover an array of topics such as food preparation, portion control, reading nutrition labels, calorie, and carbohydrate counting. Nutrition counseling is a supportive process used to set priorities, establish goals, create individualized action plans, and promote accountability for self-care (Academy of Nutrition and Dietetics, 2006). It is common practice for registered dietitians to incorporate psychology-based techniques into nutrition counseling, such as applying the transtheoretical model, motivational interviewing, and providing positive reinforcement.

In overweight and obese adults, strong evidence supports the effectiveness of multiple visits for MNT provided by a registered dietitian. According to the AND's EAL, research demonstrates improvements in anthropometric measurements as well as biochemical data. Substantial evidence supports a reduction in weight (-0.5 kg to -9.0 kg), waist circumference (-2.0 cm to -14.0 cm), BMI (-0.2 kg/m<sup>2</sup> to -7.8 kg/m<sup>2</sup>), fasting blood glucose (-5.2 mg/dL to -9.5 mg/dL), total cholesterol (-4.3 mg/dL to -59 mg/dL), LDL-cholesterol (-15 mg/dL to -47 mg/dL), HDL-cholesterol (+2.0 mg/dL to +11 mg/dL), and triglycerides (-12 mg/dL to -60 mg/dL) (Evidence Analysis Library, 2015). Despite the strong evidence to support the effectiveness of MNT, not all overweight and obese individuals have access to MNT services.

#### Educational Interventions

Nutrition education in school provides students with the knowledge to develop proper eating habits, along with necessary skills to make well-informed decisions regarding their health in the future. Students learn to choose healthy foods through effective nutrition education. Nutrition education standards are not mandatory, although school districts are encouraged to use them to develop a comprehensive kindergarten through twelfth grade nutrition education curriculum. Nutrition education standards integrate health, science, and mathematics content with real-world applications through various learning activities. When students see the connection between what they are learning and real-life examples, their motivation and learning intensifies. Key concepts presented in Wisconsin's nutrition education standards promote health literacy, food safety, food preparation, healthy eating behaviors, nutrition for growth, health, and energy. Wisconsin's nutrition education standards support variety, moderation, and balance in food choices, with the fundamental goal of engaging students in their education as they make healthier choices for themselves and their families. Schools play a significant role in helping students develop healthy eating habits by providing nutritious meals and snacks through the schools' meal programs. Additionally, school district wellness policies can implement nutrition education content by establishing healthy school environments (Wisconsin Department of Public Instruction, 2009).

Physical education in school benefits both academic performance and physical activity patterns of students. Students learn to make informed decisions and understand the value of leading an active lifestyle through effective physical education. Similar to nutrition education standards, physical education standards are not mandatory, although school districts are encouraged to use them. Physical education standards provide developmental guidance for a consistent kindergarten through twelfth grade physical education curriculum. Key concepts presented in Wisconsin's physical education standards promote physical fitness, healthy physical activity behaviors, skill development, stress reduction, improved judgement, strengthened peer

relationships, goal-setting, self-monitoring, self-discipline, self-confidence, and self-esteem. Integrating listening, speaking, reading, and writing into physical skills and activities creates cross-curricular connections, making learning relevant and meaningful to students (Wisconsin Department of Public Instruction, 2010). While nutrition and physical education are necessary components of students' overall education, standards are not mandated by schools, and support is only temporary as curriculums end before students reach an independent stage of adulthood. Behavioral Interventions

Commercial weight loss programs are popular treatment options for overweight and obese adults, although their efficacy is unclear. In 2014, Americans spent nearly \$2.5 billion on commercial weight loss services, with increases projected for years to come. Currently, Weight Watchers, Nutrisystem, and Jenny Craig are the top three programs dominating the weight loss services industry. These three programs are high intensity, with a focus on goal-setting, selfmonitoring, and group support. Nutrisystem and Jenny Craig also endorse low calorie meal replacements. Results of a systematic review indicated that Weight Watchers' participants lost more weight than control participants, which they sustained beyond 12 months. Researchers concluded that Weight Watchers has weight loss efficacy, yet it may not be superior to behavioral counseling. Weight Watchers was also the most cost-effective weight management strategy compared to other commercial programs. Nutrisystem demonstrated greater short-term weight loss compared to control and behavior counseling participants, however, long-term results were not identified. Jenny Craig participants sustained more weight loss than control and behavior counseling participants, although Jenny Craig was more expensive because it includes the price of meal replacements. Based on these findings, it may be practical for healthcare professionals to refer patients to Weight Watchers or Jenny Craig if they lack the time, training,

or ancillary staff to deliver behavioral counseling in their practices. However, high program costs may make commercial weight loss services unaffordable for many individuals (Gudzune et al., 2015).

An innovative, long-term approach to obesity prevention should address the obesogenic environments that promote sedentary lifestyles, and the consumption of highly processed, calorically-dense diets. Shaping obesogenic environments to better support healthful decisions has the potential to be a key aspect of a successful obesity prevention intervention. Thus, in order to develop effective environmental interventions in relation to obesity, we must understand how individuals and groups interact within their environments, in terms of physical activity and food intake (Townshend & Lake, 2017).

## **Technological Interventions**

Strong evidence supports the role of physical activity in managing obesity and other noncommunicable diseases. In healthcare, there is general consensus that technological interventions, including activity monitors, can potentially increase physical activity in patients. A qualitative interview-based study explored how physicians prescribed activity monitors to patients with cardiometabolic diseases. Results revealed that most physicians had never prescribed activity monitors, whereas they frequently prescribed blood glucose, blood pressure, or other self-monitoring devices. Reasons for nonprescription included a lack of interest in the data collected, a lack of evidence for data accuracy, concerns about work overload possibly resulting from automatic data transfer, and the risk of patients becoming addicted to data. Current activity monitor features are popular amongst consumers, but do not meet the needs of physicians. Physicians expected future technologies to measure physical activity intensity and duration accurately while providing understandable, motivating feedback. Understanding

physicians' expectations is a preliminary step in designing future technologies that can be widely used in clinical settings and facilitate physical activity prescription. Major healthcare stakeholders, including patients, physicians, researchers, and information technology firms should be involved in developing the most effective methods for integrating activity monitors into patient care (Bellicha, Macé, & Oppert, 2017). Moving beyond fitness tracking, activity monitors, smart watches, or their successors could provide an early warning of disease, aid in diagnosis and treatment, and contribute to a deeper understanding of human health (Savage, 2017).

In recent years, several companies have emerged as leaders of the activity monitor industry. Apple, Fitbit, Garmin, Jawbone, Microsoft, Misfit, Moov, Nike, Polar, Samsung, Striiv, TomTom, and Withings are companies at the forefront of the activity monitor industry. Activity monitors differ by price, size, style, battery life, compatibility, and tracking features. Wearable devices currently on the market range from approximately \$20 to \$600. A supplemental report will summarize, compare, and display product details of 10 current, bestselling, commercially-available activity monitors.

Beyond immediate health feedback, activity monitors also provide immediate reinforcement. Many activity monitors allow individuals to set short-term and long-term goals and compare their self-monitoring data to their goals, which supports self-efficacy. When individuals reach their goals, they may receive virtual badges, phone notifications, or email messages as positive reinforcement, which are intended to be self-motivating (Ross & Wing, 2016). Activity monitors are typically used for personal health or behavioral weight management programs, but are rarely tested as intervention tools. Below is a discussion of the

current evidence on the validity and utility of activity monitors to improve health outcomes in adults.

#### **Activity Monitor Validity**

A validity study by Shcherbina et al. assessed the accuracy of several commercially available wrist-worn devices in estimating heart rate and energy expenditure (2017). Seven commercially-available wrist-worn devices (Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2) were evaluated during varying intervals of physical activity. Participants were selected based on physical characteristics, including age, height, weight, wrist circumference, BMI, skin tone, and fitness level. Sixty healthy, demographically diverse adults ranging from 21 to 64 years of age participated in this study. Participants were from the Stanford, California area, and were distributed equally with 29 males and 31 females. While performing the standardized exercise protocol, participants wore up to four devices and simultaneously underwent continuous electrocardiographic monitoring and indirect calorimetry as FDA-approved gold standard measurements of comparison. The exercise protocol involved five-minute intervals of sitting, walking, fast walking, running, fast running, cycling, and intense cycling. All 60 participants completed the study (Shcherbina et al., 2017).

Results indicated that under laboratory-controlled conditions, six of the wrist-worn devices reported heart rate within five percent error range, while the Samsung Gear S2 achieved a 5.1 percent error rate when measuring heart rate. Error in estimation of energy expenditure was considerably higher than for heart rate for all devices. None of the wrist-worn devices reported energy expenditure within 20 percent error range. Energy expenditure error rates varied from 24 percent for the Fitbit Surge to 97.7 percent for the PulseOn. Researchers concluded that in a diverse group of individuals, heart rate measurements were within acceptable error range,

while energy expenditure estimates were not. Of the seven devices tested, the Apple Watch had the most favorable error profile, while the Samsung Gear S2 had the least favorable error profile. Strengths of this study were the highly diverse sample of participants with different ages, BMIs, and skin tones, the use of numerous consumer and gold standard comparison devices, the standardized exercise protocol to stimulate low and high intensity, and examining several different activity domains collected by the devices. Although validating wrist-worn devices in a laboratory-controlled setting was a strong starting point, it was also a limitation of this study because the results cannot be generalized to free-living conditions as they are intended for. The findings of this study were consistent with previous validity studies. Wallen, Gomersall, Keating, Wisloff, and Coombes assessed the accuracy of heart rate watches, in which heart rate error was within one to nine percent of reference standards, while energy expenditure estimates differed by 43 percent from reference standards (2016). Findings of this study add to the literature on wearable devices by developing error models and proposing a clinical standard for acceptable error. The validation data is important for researchers, as well as consumers and practitioners interested in the clinical application of wrist-worn activity monitors (Shcherbina et al., 2017).

Ferguson et al. assessed the concurrent validity of a selection of consumer-grade accelerometer-based activity monitors against two research-grade multi-sensor accelerometers in free-living conditions (2015). Seven activity monitors, including the Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike Fuelband, Striiv Smart Pedometer, and Withings Pulse were compared to two research-grade devices, the BodyMedia SenseWear and ActiGraph GT3X+ over a period of 48 hours. Twenty-one healthy adults aged 20 to 59 years, comprised of 10 males and 11 females from South Australia participated in this study. Demographic data

including date of birth, gender, height, weight, and dominant hand were obtained from all participants for device initialization and calibration. All nine devices were fitted to the participant in the following locations: BodyMedia SenseWear on the left upper arm; Nike Fuelband, Jawbone UP, and Misfit Shine on the left wrist; and ActiGraph GT3X+, Fitbit One, Fitbit Zip, Withings Pulse, and Striiv Smart Pedometer on the right side of the waist on an elasticized belt. Participants were instructed to wear all nine activity monitors simultaneously. To ensure free-living conditions were represented, the wear period was not limited to weekdays or weekends, and guidelines were not provided regarding physical activity or sleep. Over the 48hour wear period, step counting, moderate-to-vigorous intensity physical activity, sleep duration, and total daily energy expenditure data was collected and quantified. All 21 participants completed the study (Ferguson et al., 2015).

Results indicated that all consumer-grade activity monitors showed strong validity for the measurement of steps (r = 0.94 - 0.99) and sleep duration (r = 0.82 - 0.92), and moderate validity for the measurement of moderate-to-vigorous intensity physical activity (r = 0.52 - 0.91) and total daily energy expenditure (r = 0.74 - 0.81). However, the validity of the devices varied considerably within each activity construct. For example, the Misfit Shine undercounted the measurement of moderate-to-vigorous intensity physical activity (mean = 53.3 minutes) compared to the ActiGraph GT3X+ (mean = 58.5 minutes), while the Striiv Smart Pedometer overcounted the measurement of moderate-to-vigorous intensity physical activity (mean = 249 minutes) compared to the ActiGraph GT3X+. Researchers concluded that in free-living conditions, the Fitbit One, Fitbit Zip, and Withings Pulse were the strongest performers compared to the reference devices. Strengths of this study include the use of numerous consumer and reference devices, testing the devices in free-living conditions as they are designed

for, and examining several different activity variables collected by the devices. Limitations of this study were the small sample size, and varying validity if activity monitors are worn in locations other than the hip or wrist. The findings of this study were consistent with previous validity studies, which have similarly found Fitbit activity monitors to be highly valid for measuring step counts in healthy subjects. Although, the scientific evaluation of these devices is a challenge due to the rapidly evolving activity monitor market (Ferguson et al., 2015).

## **Activity Monitor Utility**

A randomized controlled trial by Cadmus-Bertram, Marcus, Patterson, Parker, and Morey evaluated the feasibility and efficacy of integrating a Fitbit tracker and website into a physical activity intervention for postmenopausal women (2015). Fifty-one overweight or obese, inactive, postmenopausal women from the San Diego, California area participated in this study. Participants attended three appointments to receive baseline assessments, orientation, and final assessments. At the second visit, participants were randomized to a 16-week intervention group. Each participant received either a Fitbit One or a basic pedometer with printed materials and a goal-setting process. The Fitbit group received additional software installation and usage training, and a follow-up call after four weeks to evaluate progress. Both groups were asked to wear the Fitbit One or pedometer every day throughout the 16-week intervention period (112 prescribed days), walk 10,000 steps per day, and perform 150 minutes of moderate-to-vigorous intensity physical activity per week (Cadmus-Bertram et al., 2015).

Compared to baseline measurements, the Fitbit group significantly increased physical activity by  $789\pm1,979$  steps per day (p=0.01),  $38\pm83$  minutes of moderate-to-vigorous intensity physical activity in 10-minute bouts (p=0.008), and a total of  $62\pm108$  minutes of moderate-to-vigorous intensity physical activity per week (p<0.001). The pedometer group experienced non-

significant increases in physical activity. Feedback indicated that participants were most engaged with the Fitbit tracker, while participants were least engaged with the website. One hundred percent of women reported liking the Fitbit One, wearing it on 95 percent of intervention days. Ninety-six percent of women rated the Fitbit One as helpful, opposed to 32 percent of women who rated the pedometer as helpful. Researchers concluded that the Fitbit intervention was associated with increased steps and physical activity at 16 weeks, while no change was observed in the pedometer group. The Fitbit One was well-accepted in this sample of women, contributing to the significant increase in physical activity. The 62-minute increase of moderate-to-vigorous intensity physical activity per week observed in the Fitbit group was substantial, especially if maintained over time. Researchers believe that physical activity interventions can be strengthened by leveraging consumer technologies that align with behavior change theories. Strengths of this study include the use of baseline and final questionnaires for detailed participant feedback, use of the ActiGraph GT3X+ as a reference device during baseline and final assessments, and use of Fitbit data to corroborate adherence. Limitations include a small sample size, short intervention period, and lack of generalizability. This study's findings differ from previous activity monitor utility studies. Thompson, Kuhle, Koepp, McCrady-Spitzer, and Levine found a Fitbit with feedback did not increase physical activity among older adults (2014). The different findings may be attributed to the age range of 16-years between participants in this study (Cadmus-Bertram et al., 2015).

O'Brien et al. used a cross-sectional study design to explore the utility of a noncommercial activity monitor to characterize activity profiles in late life depression (2017). A total of 59 subjects over the age of 60 from northeast England participated in this study. Twentynine subjects fulfilled Diagnostic and Statistical Manual of Mental Disorders criteria for current

major depression. Thirty subjects of equivalent age without self-reported history of depression or current depression comprised the control group. Each participant received a wrist-worn activity monitor and underwent neuropsychological testing over the seven-day intervention period. Additionally, demographic information, current medications, mood evaluation, social functioning, quality of life, activities of daily living, physical, and mental wellbeing were assessed at baseline and day seven. Due to less than seven days of battery life, the initial wristworn activity monitor was switched with an identical, fully charged device between days two and six (O'Brien et al., 2017).

Results indicated that physical activity was significantly reduced in participants with late life depression compared to healthy controls (p<0.001). The difference in activity levels between the groups was greatest during the morning and early afternoon. Furthermore, participants with late life depression showed significantly slower fine motor movements (p<0.001), lower quality of life scores (p<0.001), and reduced activities of daily living (p<0.001)compared to healthy controls. Researchers concluded that quality of life and activities of daily living measures were strongly correlated with physical activity, while self-reported measures of loneliness and social support were not. High resolution analysis of accelerometer-derived physical activity may provide an appropriate indication of depression in older adults. Lastly, since exercise has been proposed as a treatment for individuals with depression, wearable devices may play a positive role in monitoring levels of activity when used therapeutically. Strengths of this study include the use of an unobtrusive, waterproof wrist-worn activity monitor, a cohort of currently depressed older adults, and high compliance. A limitation of this study was that causality between physical activity and other key variables could not be determined due to the cross-sectional study design. Likewise, the association between physical activity, depression,

and cognition may be interrelated. This was the first study to objectively characterize the quality of physical activity in late life depression, suggesting that when used therapeutically, wearable devices have potential to objectively monitor levels of activity (O'Brien et al., 2017).

# Discussion

Both validity studies simultaneously compared seven commercial activity monitors to research-grade devices. Ferguson et al. found Fitbit activity monitors to be of highest validity, while Shcherbina et al. found the Apple Watch to have the most favorable error profile. The selection of activity monitors was different in both studies, and not all activity monitor functions were validated. When using healthy subjects, the most accurate Fitbit measurement was step counts, while the most accurate measurement for the Apple Watch was heart rate. The environment of these studies also differed from each other; the first study was conducted in a laboratory-controlled setting, while the second study was conducted in free-living conditions. When comparing activity monitor utility studies, both concentrated on older adult populations. The third study discussed used a Fitbit One, while the fourth study used a non-commercial activity monitor. Cadmus-Bertram et al. found that a Fitbit intervention was associated with significantly increased physical activity in postmenopausal, overweight or obese women compared to a pedometer intervention. O'Brien et al. used the activity monitor not as an intervention, but to assess activity levels, and found a significant reduction in general physical activity, lower quality of life, and reduced activities of daily living in depressed older adults. The results of the fourth study discussed could lead into another experimental study utilizing a commercial activity monitor as an intervention tool to increase physical activity in depressed older adults. Although both outcomes are in favor of activity monitor usage, both studies lacked evidence of long-term health benefits.

## Conclusion

The purpose of this literature review was to critically analyze current evidence on the validity and utility of activity monitors to improve health outcomes in adults. Activity monitor validity studies suggest Fitbits and Apple Watches are high quality commercial devices. In healthy subjects, Fitbits count steps most accurately, and Apple Watches precisely measure heart rate. Utility studies suggest activity monitors are motivating devices that can increase physical activity, may support other positive health outcomes, and may be potential indicators used in the detection of disease. Future research must expand to keep up with the rapidly evolving activity monitors should be regulated by predefined boundaries of accuracy based on reference device standards (Chowdhury, Western, Nightingale, Peacock, & Thompson, 2017). Further research should evaluate which activity monitor features are most effective, examine all health outcomes associated with utilizing activity monitors, and determine which populations are most receptive to activity monitors.

Activity monitors can be useful additions to healthy lifestyle interventions. Activity monitor functions also have the potential to reach reference device standards. As research and product development progress, activity monitors may have practical clinical applications for all populations. Healthcare professionals could provide evidence-based recommendations regarding appropriate activity monitor selection. With automatic data transfer of validated activity monitors, healthcare professionals could accurately track clients' physical activity habits, heart rate, energy expenditure, sleep patterns, food and beverage consumption, and weight fluctuations. Direct access to such data would save an immense amount of time, provide a holistic picture of health for each client, and enable healthcare professionals to personalize care

more than ever before. Future technologies may be able to provide an early warning of disease, aid in diagnosis and treatment, and contribute to a deeper understanding of human health (Savage, 2017). Thus, the individualized approach to care by means of utilizing activity monitor data will result in better health outcomes for all populations, ultimately leading to a decreased prevalence of overweight and obesity, chronic illnesses, mortality, and healthcare costs.

## CHAPTER III: METHODOLOGY

The AND's EAL is an online resource synthesizing the strongest, most relevant nutritional research on important dietetic practice questions. The EAL is a series of systematic reviews developed by Academy members for Academy members. Expert Academy members use a predefined approach and criteria to document each step, ensuring objectivity, transparency, and reproducibility of the Evidence Analysis process. The EAL provides bibliographies, evidence summaries, worksheets, conclusion statements, and grades, as well as recommendations, recommendation strength and narrative, algorithms, and links to evidence. The EAL enhances the credibility of the dietetics profession by assisting dietetic practitioners in utilizing evidence-based practice (Academy of Nutrition and Dietetics, 2018). This Evidence Analysis project follows the five steps of the Evidence Analysis process, described below.

#### **Step One: Formulate the Evidence Analysis Question**

The first step in the Evidence Analysis process focuses on a specific question in a defined area of practice. High quality research questions are developed using an analytical framework to identify links between factors and outcomes. The AND's NCP should serve as the framework, while the PICO format should be used to format questions. PICO is an acronym for population, intervention, comparison, and outcome (Academy of Nutrition and Dietetics, 2016). The research questions in this Evidence Analysis project are as follows.

- 1. How valid are the functions of current, consumer-grade activity monitors compared to research-grade devices?
- 2. Does physical activity improve and weight loss occur as a result when adults utilize current, consumer-grade activity monitors?

## Step Two: Gather and Classify the Evidence

After the Evidence Analysis question is formulated, research must be gathered and classified. This step involves creating a search plan to conduct a thorough literature search. Ensuring that all relevant evidence is reviewed is one of the most important aspects of the Evidence Analysis process. If some evidence is missed, the conclusion statement may be misleading. The search plan should delineate the inclusion and exclusion criteria, key search terms or phrases, and outcomes necessary to conduct a thorough literature search. Research is classified by type of evidence, with classes differentiating between primary (Class A, B, C, and D) and secondary (Class M, R, and X) reports (Academy of Nutrition and Dietetics, 2016). Research articles that meet the predefined criteria will be included for further evaluation. Articles that do not meet the criteria will be excluded along with appropriate reasoning. This Evidence Analysis project adheres to the following search plan.

Research Question	<ol> <li>How valid are the functions of current, consumer-grade activity monitors compared to research-grade devices?</li> <li>Does physical activity improve and weight loss occur as a result when adults utilize current, consumer-grade activity monitors?</li> </ol>
Date of Literature Review	September 2018
Inclusion Criteria	<ul> <li>Language: English</li> <li>Research: primary (Class A, B, C, and D)</li> <li>Year Range: 2014 to 2018</li> <li>Participant Age Range: adults 18 to 80 years</li> <li>Participant Health Status: healthy or with one health condition related to diet and/or physical inactivity</li> </ul>
Exclusion Criteria	<ul> <li>Language: all languages that are not English</li> <li>Research: secondary (Class M, R, and X)</li> <li>Year Range: prior to 2014</li> <li>Participant Age Range: children less than 18 years and adults greater than 80 years</li> <li>Participant Health Status: with more than one health condition that is related or unrelated to diet and/or physical inactivity</li> </ul>

Search Terms/Phrases	٠	Activity monitor
	٠	Fitness tracker
Electronic Databases	٠	PubMed
Included Articles	•	Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. <i>PLoS ONE</i> , <i>13</i> (2), e0192691. doi:10.1371/journal.pone.0192691
	•	Cadmus-Bertram, L., Gangnon, K., Wirkus, E. J., Thraen- Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. <i>Annals of Internal Medicine</i> , <i>166</i> (8), 610– 612. doi:10.7326/L16-0353
	•	Cadmus-Bertram, L., Marcus, B., Patterson, R., Parker, B., & Morey, B. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. <i>American Journal</i> <i>of Preventive Medicine</i> , <i>49</i> (3), 414–418. doi:10.1016/j.amepre.2015.01.020
	•	Chum, J., Kim, M., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., Samaan, Z. (2017). Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. <i>Evidence-Based Mental Health</i> , <i>20</i> (4), 128–133. doi:10.1136/eb-2017-102763
	•	Cook, J., Prairie, M., & Plante, D. (2017). Utility of the Fitbit Flex to Evaluate Sleep in Major Depressive Disorder: A comparison against polysomnography and wrist-worn actigraphy. <i>Journal of Affective Disorders</i> , <i>217</i> , 299–305. doi: 10.1016/j.jad.2017.04.030
	•	Ferguson, T., Rowlands, A., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. <i>International Journal of Behavioral Nutrition and Physical</i> <i>Activity, 12</i> , 42. doi:10.1186/s12966-015-0201-9
	•	Gomersall, S., Ng, N., Burton, N., Pavey, T., Gilson, N., & Brown, W. (2016). Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. <i>Journal of Medical Internet</i> <i>Research</i> , <i>18</i> (9), e239. doi:10.2196/jmir.5531
	•	Gualtieri, L., Rosenbluth, S., & Phillips, J. (2016). Can a Free Wearable Activity Tracker Change Behavior? The Impact of Trackers on Adults in a Physician-Led Wellness Group. <i>JMIR Research Protocols</i> , <i>5</i> (4), e237. doi:10.2196/resprot.6534
	•	Naslund, J., Aschbrenner, K., Scherer, E., McHugo, G., Marsch, L., & Bartels, S. (2016). Wearable Devices and Mobile Technologies for Supporting Behavioral Weight Loss

		Among People with Serious Mental Illness Psychiatry	
		Research 244 139–144 doi:10.1016/j.psychres.2016.06.056	
	•	Mahar C Ryan I Ambrosi C & Ednay S (2017) Users'	
	•	avpariances of waarable activity trackers: a cross sectional	
		study DMC Dublic Health 17 990 doi:10.1196/s12990	
		Study. DMC Fublic Health, 17, 880. doi:10.1180/812889-	
		01/-4888-1	
	•	Rosenberger, M., Buman, M., Haskell, W., McConnell, M.,	
		& Carstensen, L. (2016). 24 Hours of Sleep, Sedentary	
		Behavior, and Physical Activity with Nine Wearable	
		Devices. Medicine and Science in Sports and Exercise,	
		48(3), 457–465. doi:10.1249/MSS.0000000000000/78	
	•	Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H.,	
		Christle, J., Hastie, T., Ashley, E. (2017). Accuracy in	
		Wrist-Worn, Sensor-Based Measurements of Heart Rate and	
		Energy Expenditure in a Diverse Cohort. Journal of	
		<i>Personalized Medicine</i> , 7(2), 3. doi:10.3390/jpm7020003	
Excluded Articles with	•	Abrantes, A., Blevins, C., Battle, C., Read, J., Gordon, A., &	
Reason		Stein, M. (2017). Developing a Fitbit-Supported Lifestyle	
		Physical Activity Intervention for Depressed Alcohol	
		Dependent Women. Journal of Substance Abuse Treatment,	
		80, 88–97. doi:10.1016/j.jsat.2017.07.006	
		• Participants used two consumer-grade activity monitors	
		inconsistently	
		• Assessed utility of activity monitors in participants with	
		more than one health condition	
	٠	Berendsen, B., Hendriks, M., Meijer, K., Plasqui, G.,	
		Schaper, N., & Savelberg, H. (2014). Which activity monitor	
		to use? Validity, reproducibility and user friendliness of three	
		activity monitors. BMC Public Health, 14, 749.	
		doi:10.1186/1471-2458-14-749	
		• Participants did not use a consumer-grade activity	
		monitor	
	•	Cochrane, S., Chen, S., Fitzgerald, J., Dodson, J., Fielding,	
		R., King, A., Kaplan, R. (2017). Association of	
		Accelerometry-Measured Physical Activity and	
		Cardiovascular Events in Mobility-Limited Older Adults:	
		The LIFE (Lifestyle Interventions and Independence for	
		Elders) Study. Journal of the American Heart Association:	
		<i>Cardiovascular and Cerebrovascular Disease</i> , <i>6</i> (12),	
		e00/215. doi:10.1161/JAHA.117.007215	
		• Participants did not use a consumer-grade activity	
		monitor	
	•	Correa, J., Apolzan, J., Shepard, D., Heil, D., Rood, J., &	
		Martin, C. (2016). Evaluation of the ability of three physical	
		activity monitors to predict weight change and estimate	

	energy expenditure. Applied Physiology, Nutrition, and
	Metabolism, 41(7), 758–766. doi:10.1139/apnm-2015-0461
	<ul> <li>Assessed predictability of body weight changes</li> </ul>
	• LaMonte, M., Lewis, C., Buchner, D., Evenson, K.,
	Rillamas-Sun, E., Di, C., Shumaker, S. (2017). Both Light
	Intensity and Moderate-to-Vigorous Physical Activity
Measured by Accelerometry Are Favorably Assoc	
Cardiometabolic Risk Factors in Older Women: The	
Objective Physical Activity and Cardiovascular He	
	(OPACH) Study. Journal of the American Heart
	Association: Cardiovascular and Cerebrovascular Disease,
	6(10), e007064. doi:10.1161/JAHA.117.007064
	<ul> <li>Participants did not use a consumer-grade activity</li> </ul>
monitor	
	• O'Brien, J., Gallagher, P., Stow, D., Hammerla, N., Ploetz,
	T., Firbank, M., Olivier, P. (2017). A study of wrist-worn
	activity measurement as a potential real-world biomarker for
	late-life depression. <i>Psychological Medicine</i> , 47(1), 93-102.
	doi:10.1017/S0033291716002166
	• Participants did not use a consumer-grade activity
	monitor
Summary of Articles	Number of Primary Research Articles Identified: 18
Identified to Review	Number of Primary Research Articles Excluded: 6
	Total Number of Primary Research Articles Included: 12

# Step Three: Critically Appraise Each Article

Research articles that meet the inclusion criteria are evaluated for methodologic quality in this step. Articles are appraised individually based on the appropriateness of the study design and the quality of how the study was conducted by using the AND's Quality Criteria Checklist (QCC) and EAL worksheets. Key information is abstracted and entered into the worksheets. The QCC worksheet assesses risk of bias by asking questions related to relevance and validity, while determining if the strength of evidence overall is rated as negative, neutral, or positive quality. The EAL worksheet summarizes the methodology, results, authors' conclusion, and reviewer's comments (Academy of Nutrition and Dietetics, 2016). Studies with negative quality

evidence were not included or used in the development of either conclusion statement in this Evidence Analysis project.

#### **Step Four: Summarize the Evidence**

After the research has been critically appraised, evidence from each article is summarized and displayed. This step involves the creation of an overview table and an evidence summary. The overview table includes key information such as study designs, sample sizes, interventions, outcomes, and quality ratings, as well as citations for the included research articles. Data for the overview table is transferred from the QCC and EAL worksheets. The overview table is beneficial because it allows healthcare professionals to visually compare the studies. The evidence summary is a concise description of the overall findings of the Evidence Analysis process (Academy of Nutrition and Dietetics, 2016). Activity monitor validity and utility data from neutral and positive quality studies has been summarized and displayed in an overview table and evidence summary.

## Step Five: Write and Grade the Conclusion Statement

The final step in the Evidence Analysis process involves establishing a conclusion statement to answer the research question, along with assigning a grade to the conclusion statement. The grade reflects the overall strength of the available supporting evidence in forming the conclusion statement. The grading scale used by the AND includes: Grade I (good/strong), II (fair), III (limited/weak), IV (expert opinion only), and V (not assignable) (Academy of Nutrition and Dietetics, 2016). Conclusion statements for both research questions in this Evidence Analysis project have been established based on the available supporting evidence. The overall strength of the supporting evidence has also been graded, ranging from Grade I to V.

# Supplemental Comparison Report

Following the completion of the Evidence Analysis process, product details of 10 current, best-selling, commercially-available activity monitors will be summarized and displayed in a supplemental comparison report to fulfill objective C of this Evidence Analysis project. Data has been collected exclusively from product manufacturer websites to maintain the EAL's standards of objectivity, transparency, and reproducibility. If evidence is deemed adequate and conclusive, dietitians and other healthcare professionals may draft evidence-based guideline recommendations regarding appropriate activity monitor selection.
# CHAPTER IV: RESULTS

Using the PubMed database, a total of 18 studies were identified relating to activity monitor validity and utility in adults. Six studies were excluded because they assessed participants with more than one health condition, the predictability of body weight changes, or activity monitors that were not consumer-grade devices. Twelve studies were included for further evaluation. Of the 12 studies, seven were validity studies and five were randomized controlled trials, non-randomized crossover trials, and a cross sectional study relating to activity monitor utility. An overview table summarizing and displaying key information from each study can be found in Appendix B. Below are brief summaries of each study, followed by the graded conclusion statements that answer both research questions.

#### Benedetto et al., (2018) (Quality Rating: +)

Benedetto et al., (2018) assessed in a controlled, research environment the accuracy and precision of the Fitbit Charge 2 for measuring heart rate with respect to the ProComp Infiniti T7500M, a gold standard electrocardiograph. The Fitbit Charge 2 exhibited a mean bias of -5.9 beats per minute (95% CI). The limits of agreement, which indicate the precision of individual measurements, between the Fitbit Charge 2 and ProComp Infiniti T7500M were wide. The upper limit of agreement was +16.8 beats per minute, whereas the lower limit of agreement was -28.5 beats per minute. The intraclass correlation coefficient, used as an alternative measure of agreement between the Fitbit Charge 2 and ProComp Infiniti T7500M was 0.21 (95% CI). Researchers concluded that the Fitbit Charge 2 tends to underestimate heart rate with moderate bias on average, although precision is poor for individual measurements, which could be underestimated by as much as 30 beats per minute.

# Cadmus-Bertram et al., (2017) (Quality Rating: ø)

Cadmus-Bertram et al., (2017) determined the heart rate accuracy measured by four commercial, light-emitting diode-dependent, wrist-worn activity trackers. When participants rested, the limits of agreement were best for the Fitbit Surge (-5.1 to 4.5 beats per minute) and worst for the Basis Peak (-17.1 to 22.6 beats per minute). When participants exercised at 65 percent of their maximum heart rate, the limits of agreement were relatively poor for all the activity trackers, ranging from -22.5 to 26.0 beats per minute for the Mio Fuse to -41.0 to 36.0 beats per minute for the Fitbit Charge. At rest, the repeatability coefficient ranged from 4.2 beats per minute for the Fitbit Surge to 19.3 beats per minute for the Basis Peak to 23.7 beats per minute for the Mio Fuse. Researchers concluded that all activity trackers were more accurate at rest than during moderate exercise, implying that more heart rate feature research is needed.

# Cadmus-Bertram et al., (2015) (Quality Rating: +)

Cadmus-Bertram et al., (2015) evaluated within a randomized controlled trial, the feasibility and preliminary efficacy of integrating a Fitbit tracker and website into a physical activity intervention for postmenopausal, overweight or obese women. After the 16-week intervention, the Fitbit group significantly increased physical activity by 789±1,979 steps per day (p=0.01), moderate-to-vigorous intensity physical activity in 10-minute bouts by 38±83 minutes per week (p=0.008), and moderate-to-vigorous intensity physical activity by 62±108 minutes per week (p<0.001), compared to non-significant increases in the pedometer group (between-group p-values were 0.11, 0.28, and 0.30, respectively). The Fitbit group wore the Fitbit One on 95 percent of intervention days, 96 percent of women reported liking the website, and 100 percent

of women reported liking the Fitbit One. Although the study had confounding variables, researchers concluded that the Fitbit One was well-accepted in this sample of women, was associated with increased steps and physical activity at 16 weeks, and that physical activity interventions can be strengthened by leveraging consumer technologies that align with behavior change theories.

## Chum et al., (2017) (Quality Rating: ø)

Chum et al., (2017) aimed to understand patients' perceived benefit from the Fitbit One and explore themes associated with patient experiences, as well as compare the perceived benefit, patient factors, Fitbit usage, and Beck's Depression Inventory scores. Of the 36 patients who underwent the BRAVE study and completed interviews, 23 patients found the Fitbit One to be helpful for their physical activity. Themes of positive experiences included self-awareness, peer motivation, and goal-setting opportunities. Themes of negative experiences included inconvenience, inaccuracies, discouragement, and disinterest. There was a significant relationship between total Fitbit One usage and perceived benefit. The mean number of weeks of Fitbit One use for those who found the Fitbit helpful was 18.57 and 12.27 weeks for those who did not (p<0.001). Interestingly, there was no significant relationship between perceived benefit of the Fitbit One and percent change in Beck's Depression Inventory scores, which contradicts previous literature supporting Fitbit use in treating depression. Researchers concluded that the Fitbit One may be useful for patients with varying characteristics, although strengths and limitations of activity trackers should be considered when implementing them to motivate patients with depression.

Cook et al., (2017) (Quality Rating: +)

Cook et al., (2017) evaluated the utility of the Fitbit Flex to estimate sleep in a wellcharacterized cohort of adult patients with major depressive disorder relative to gold standard polysomnography and validated actigraphy (Actiwatch-2; AW-2). Compared to polysomnography, the Fitbit Flex significantly overestimated total sleep time (mean difference of 46.0 minutes, p<0.0001) and sleep efficiency (mean difference of 8.1%, p<0.0001), significantly underestimated wake after sleep onset (mean difference of -44.0 minutes, p<0.0001), while sleep onset latency was quite similar (mean difference of -2.0 minutes, p=0.72). The Fitbit Flex demonstrated high sensitivity and accuracy with low specificity. Researchers concluded that the Fitbit Flex is not an adequate substitute for polysomnography when quantifying sleep in adults with major depressive disorder. However, the Fitbit Flex does demonstrate similar performance characteristics to a standard actigraph.

# Ferguson et al., (2015) (Quality Rating: +)

Ferguson et al., (2015) assessed the concurrent validity of a selection of consumer-grade, accelerometer-based activity monitors compared to two research-grade accelerometers in freeliving conditions. All activity monitors measured steps, and correlations with reference devices were very strong (r=0.94-0.99). Five activity monitors measured moderate-to-vigorous intensity physical activity, and correlations ranged from moderate to strong (r=0.52-0.91). Four activity monitors measured sleep, and all correlated strongly with the reference device (r=0.82-0.92). Five activity monitors measured total daily energy expenditure, and correlations were moderate to strong (r=0.74-0.81). Researchers concluded that the Fitbit One, Fitbit Zip, and Withings Pulse were the strongest performers.

# Gomersall et al., (2016) (Quality Rating: ø)

Gomersall et al., (2016) compared Fitbit One and Jawbone UP estimates of steps, moderate-to-vigorous intensity physical activity, and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context. Correlations for steps and moderateto-vigorous intensity physical activity were strong for both devices, although higher for the Fitbit One (r=0.85 for steps and p=0.80 for moderate-to-vigorous intensity physical activity) than for the Jawbone UP (r=0.75 for steps and p=0.75 for moderate-to-vigorous intensity physical activity). The correlation between the Jawbone UP longest idle time and ActiGraph longest sedentary bout was weak (p=0.19). Agreement between the Fitbit One and ActiGraph for the classification of active versus inactive time was substantial (k=0.68, p<0.001), while agreement between the Jawbone UP and ActiGraph was moderate (k=0.52, p<0.001). Due to modest accuracy and systematic bias, researchers concluded that both activity trackers are better suited as self-monitoring tools for consumers or in behavior change interventions rather than for the evaluation of research outcomes.

#### Gualtieri et al., (2016) (Quality Rating: ø)

Gualtieri et al., (2016) investigated the use of wearable activity trackers by adults with chronic medical conditions who have never used trackers previously. Specifically, the researchers aimed to determine (1) if participants would accept and use activity trackers to increase their physical activity; (2) if there were barriers to use besides cost and training; (3) if activity trackers would educate participants on their activity levels and support behavior change; and (4) if clinical outcomes would show improvements in participants' health. Improvements were seen in clinical outcomes, physical activity behaviors, and attitudes towards the Withings Pulse after the 12-week study. Participants lost an average of 0.5 pounds per week with a mean total weight loss of 5.97 pounds (p=0.004). Other clinical outcomes included a 9.2 percent

decrease in LDL levels (p=0.038), while changes in blood pressure were non-significant. All participants reported an increase in well-being, health education, physical activity, and confidence in their ability to lead more active lives. Researchers concluded that adding activity trackers to wellness groups can support education and behavior change, adding that it may be cost-effective to provide free or heavily subsidized activity trackers that lower the risk of chronic conditions compared to the healthcare costs required to treat illnesses after they develop.

## Maher et al., (2017) (Quality Rating: +)

Maher et al., (2017) explored users' experiences of activity trackers, including the ease of use, patterns of use, barriers to use, and perceived usefulness for tracking and modifying lifestyle behaviors, such as physical activity, diet, and sleep. The most commonly used brand of activity tracker was Fitbit (67.5%), followed by Garmin (16.5%), Apple (3.4%), Jawbone (2.5%), Samsung (1.7%), Polar (1.3%), and other (7.1%). Participants agreed that various features on their activity trackers were useful, including step counts (95%), active minutes (76%), sleep (66%), heart rate (63%), elevation or stairs climbed (58%), and calories burned (57%), while fewer participants agreed that the food intake feature was useful (36%). Overall, 94 percent of current users and 65 percent of former users agreed that they had a positive experience with their activity tracker. Researchers concluded that in general, activity trackers are used for a substantial period of time, are viewed positively by users, and are useful tools for intervening on physical activity.

## Naslund et al., (2016) (Quality Rating: +)

Naslund et al., (2016) examined whether average daily step counts measured using Fitbit Zip wearable devices was associated with weight loss and improved fitness among individuals

with serious mental illness enrolled in a six-month lifestyle program. Participants accumulated an average of 4,453.5 steps per day, with average daily step counts ranging from 1,037.6 to 11,366.3 steps. There was a significant association between participants' average daily step counts and weight loss. For every 1,000-step increase, participants experienced a decrease in weight of 1.78 pounds (p=0.0314). The relationship between average daily step counts and change in fitness was non-significant. Every 1,000-step increase corresponded to an increase of 18.79 feet on the Six-Minute Walk Test (p=0.176). Researchers concluded that wearable devices and their associated smartphone applications may serve as valuable tools for supporting weight loss efforts in individuals with serious mental illness. Additionally, providing participants with the recommendation to maintain a high average daily step count throughout participation in a lifestyle intervention may contribute to greater weight loss.

# Rosenberger et al., (2016) (Quality Rating: ø)

Rosenberger et al., (2016) compared the output from commercially available wearable devices using current standards for objective measurement of sleep, sedentary behavior, lightintensity physical activity, moderate-to-vigorous intensity physical activity, and steps in a freeliving environment. Mean error analyses for the devices ranged from 8.1 percent for the Actigraph GT3X+ to 16.9 percent for the GENEactiv when measuring sleep duration; 9.5 percent from the LUMOback to 65.8 percent for the GENEactiv when measuring sedentary behavior; 19.7 percent from the GENEactiv to 28.0 percent for the Fitbit One when measuring light-intensity physical activity; 51.8 percent from the Jawbone Up to 92.0 percent for the Nike Fuelband when measuring moderate-to-vigorous intensity physical activity; and 14.1 percent from the Actigraph GT3X+ to 29.9 percent for the Nike Fuelband when measuring total steps per day. Equivalence analyses indicated only one comparison device, the LUMOback was

significantly equivalent to standards when measuring sedentary behavior (90% CI). Researchers concluded that none of the commercial wearable devices provide all measures of the 24-hour activity model, which is currently only possible with research-grade devices, suggesting that the future of activity measurement should aim for accurate 24-hour measurement as a goal.

# Shcherbina et al., (2017) (Quality Rating: +)

Shcherbina et al., (2017) assessed under controlled laboratory conditions the accuracy of seven commercially available wrist-worn devices in estimating heart rate and energy expenditure. The lowest error in measuring heart rate (1.8%) was observed for the cycling stage (0.9%-2.7%, 95% CI), while the highest error in measuring heart rate (5.5%) was observed for the walking stage (3.9%-7.1%, 95% CI). Error in estimation of energy expenditure was considerably higher than for heart rate for all devices. Median error rates across activities varied from 27.4 percent (24.0%-30.8%, 95% CI) for the Fitbit Surge to 92.6 percent (87.5%-97.7%, 95% CI) for the PulseOn. Researchers concluded that most wrist-worn devices reported heart rate within acceptable error range (5%), while none of the wrist-worn devices reported energy expenditure within an acceptable error range. The Apple Watch had the most favorable error profile and the Samsung Gear S2 had the least favorable error profile.

# **Research Question**

1. Current, consumer-grade activity monitor functions will be within 25 percent error range of research-grade devices

# **Conclusion Statement**

Current, consumer-grade activity monitors exhibit moderate validity on average, tend to estimate step counts accurately, underestimate heart rate and energy expenditure, overestimate time asleep, and are more accurate at rest than during activity.

# Grade: II, Fair

This is a Grade II, Fair conclusion because the evidence consists of results from studies that answered the research question addressed, although there is uncertainty about generalizability, bias, research design flaws, and adequacy of sample sizes.

# **Research Question**

2. Does physical activity improve and weight loss occur as a result when adults utilize current, consumer-grade activity monitors?

# **Conclusion Statement**

Adults who utilize current, consumer-grade activity monitors as combined interventions may experience a clinically meaningful increase in steps, physical activity, and weight loss.

# Grade: II, Fair

This is a Grade II, Fair conclusion because the evidence consists of results from studies that answered the research question addressed, although there is uncertainty about generalizability, bias, research design flaws, and adequacy of sample sizes.

# CHAPTER V: CONCLUSION

# **Evidence Summary**

Current, consumer-grade activity monitors exhibit moderate validity on average and provide more precise measurements at rest compared to during activity. Activity monitors tend to estimate step counts accurately, underestimate heart rate and energy expenditure, and overestimate time asleep. The Apple Watch, Fitbits, and Withings Pulse were found to be the most valid, while the Basis Peak, Jawbone UP, Nike Fuelband, PulseOn, and Samsung Gear S2 were found to be the least valid consumer-grade devices. Adults who utilize activity monitors as combined interventions may experience a clinically meaningful increase in steps, physical activity, and weight loss. Overall, activity monitors are recognized as useful additions to healthy lifestyle interventions.

Of the seven validity studies, five assessed activity monitor accuracy in healthy subjects, one evaluated participants with major depressive disorder, and one did not specify the study population beyond reporting that participants were adults. Of the five studies exploring activity monitor utility, two were randomized controlled trials, two were non-randomized crossover trials, and one was a cross sectional study. Populations included in the activity monitor utility studies were postmenopausal women, participants with overweight or obesity, one chronic medical condition, major depressive disorder, serious mental illness, and were current or former activity tracker users. Not all studies analyzed activity monitor validity or utility in improving health outcomes in adults within the same parameters. Four validity studies were conducted under controlled, research environments utilizing gold standard comparisons, while three validity studies were conducted in free-living environments as activity monitors are intended for, utilizing common field-based devices as a comparison. Utility studies did not evaluate activity

monitors as independent interventions. Instead, most utility studies incorporated activity monitors as self-motivating, self-monitoring tools integrated into behavioral interventions with positive outcomes.

# Conclusion

Research hypotheses were relatively similar to the conclusion statements developed for this Evidence Analysis project. Activity monitor utility studies revealed that adults did in fact experience an increase in physical activity and a decrease in weight as a result. However, exact error profiles could not be determined due to the varied statistical analyses and methods of reporting results by activity monitor validity studies. Interestingly, Chum et al. found no significant relationship between perceived benefit of the Fitbit One and percent change in Beck's Depression Inventory scores, which contradicts previous literature supporting activity monitor use in treating depression (2017). Results from this Evidence Analysis project demonstrate that activity monitors are self-motivating, self-monitoring devices, and their usefulness is not dependent on their validity. Participants were motivated by activity monitors regardless of their accuracy. Although, for research purposes, activity monitors could be more effective if they were more accurate.

#### **Applications to Practice**

Activity monitors have current and future applications to practice. Currently, activity monitors create awareness by assisting individuals in tracking their weight-related behaviors. Personal behaviors play a dominant role in preventing and treating noncommunicable diseases, such as obesity, diabetes, cardiovascular disease, and certain cancers. As self-motivating, selfmonitoring devices, activity monitors may be used as intervention tools to improve adherence and support positive behavior change. Dietitians and other healthcare professionals may

encourage the use of activity monitors as a cost-effective method to implement healthy lifestyle changes, which in turn would decrease morbidity, mortality, and healthcare costs.

Activity monitors have the potential to reach reference device standards, which in turn may result in further clinical applications for all populations. Dietitians and other healthcare professionals could provide evidence-based recommendations regarding appropriate activity monitor selection. With automatic data transfer of validated activity monitors, healthcare professionals could accurately track their clients' physical activity habits, heart rate, energy expenditure, sleep patterns, food and beverage consumption, and weight fluctuations. Direct access to such data would save an immense amount of time, provide a holistic picture of health for each client, and enable healthcare professionals to personalize care more than ever before. Future wearable technologies may be able to provide an early warning of disease, aid in diagnosis and treatment, and contribute to a deeper understanding of human health (Savage, 2017). Hopefully, the individualized approach to care by means of utilizing validated activity monitor data will result in better health outcomes for all populations, ultimately leading to a decreased prevalence of overweight and obesity, chronic illnesses, mortality, and healthcare costs.

#### **Recommendations for Future Research**

It may be speculated that current activity monitor hardware and software used to estimate activity parameters, such as step counts, elevation or stairs climbed, distance traveled, heart rate, calories burned, active time, and time asleep lack proper sophistication. Product developers should consult activity monitor users, healthcare professionals, researchers, and information technology firms to gain an in-depth understanding of current activity monitor performance,

effective methods for integrating activity monitors into clinical care, and expectations for activity monitor successors.

Future research needs to expand in this area. First, research should evaluate which activity monitor features are most effective, determine which populations are most receptive to activity monitors, and examine all health outcomes associated with utilizing activity monitors. Second, research must be more comprehensive with larger sample sizes, demographically diverse subjects, stronger study designs, and longer study durations to determine long-term effects. Third, research must be consistent, if not continuous in order to keep up with the rapidly evolving consumer-grade activity monitor market. To achieve this, researchers could collaborate in the creation of an online forum to update and share evidence objectively. Evidence should also be made public in an easily readable and understandable format to encourage transparency from activity monitor manufacturers. As a final effort to improve the quality of activity monitors and associated health outcomes, predefined boundaries of accuracy based on reference device standards should be regulated.

#### BIBLIOGRAPHY

- Academy of Nutrition and Dietetics. (2018). About the Evidence Analysis Library. Retrieved from https://www.andeal.org/about
- Academy of Nutrition and Dietetics. (2017). Definition of Terms List. Retrieved from <u>https://www.eatrightpro.org/-/media/eatrightpro-files/practice/scope-standards-of-</u> practice/academydefinitionoftermslist.pdf
- Academy of Nutrition and Dietetics, Evidence Analysis Library. (2015). MNT: Weight Management. Retrieved from <u>https://www.andeal.org/topic.cfm?menu=5284&cat=5230</u>
- Academy of Nutrition and Dietetics. (2016). Evidence Analysis Manual: Steps in the Academy Evidence Analysis Process. Retrieved from

https://www.andeal.org/vault/2440/web/files/2016\_April\_EA\_Manual.pdf

- Academy of Nutrition and Dietetics. (2006). MNT Versus Nutrition Education. Retrieved from <a href="https://www.eatrightpro.org/payment/coding-and-billing/mnt-vs-nutrition-education">https://www.eatrightpro.org/payment/coding-and-billing/mnt-vs-nutrition-education</a>
- Bellicha, A., Macé, S., & Oppert, J. M. (2017). Prescribing of Electronic Activity Monitors in Cardiometabolic Diseases: Qualitative Interview-Based Study. *Journal of medical Internet research*, 19(9), e328. doi:10.2196/jmir.8107
- Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., & Actis, P. (2018).
  Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS ONE*, *13*(2), e0192691.
  doi:10.1371/journal.pone.0192691
- Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by Some Wrist-Worn
   Activity Trackers. *Annals of Internal Medicine*, *166*(8), 610–612. doi:10.7326/L16-0353

Cadmus-Bertram, L., Marcus, B., Patterson, R., Parker, B., & Morey, B. (2015). Randomized

Trial of a Fitbit-Based Physical Activity Intervention for Women. *American Journal of Preventive Medicine*, 49(3), 414–418. doi:10.1016/j.amepre.2015.01.020

- Chowdhury, E., Western, M., Nightingale, T., Peacock, O., & Thompson, D. (2017). Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PLoS ONE*, *12*(2), e0171720.
  doi:10.1371/journal.pone.0171720
- Chum, J., Kim, M., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., ... Samaan, Z. (2017).
   Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. *Evidence-Based Mental Health*, 20(4), 128–133. doi:10.1136/eb-2017-102763
- Cook, J., Prairie, M., & Plante, D. (2017). Utility of the Fitbit Flex to Evaluate Sleep in Major
   Depressive Disorder: A comparison against polysomnography and wrist-worn actigraphy.
   *Journal of Affective Disorders*, 217, 299–305. doi: 10.1016/j.jad.2017.04.030
- Ferguson, T., Rowlands, A., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity, 12*, 42. doi:10.1186/s12966-015-0201-9
- Gomersall, S., Ng, N., Burton, N., Pavey, T., Gilson, N., & Brown, W. (2016). Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. *Journal of Medical Internet Research*, 18(9), e239. doi:10.2196/jmir.5531
- Gualtieri, L., Rosenbluth, S., & Phillips, J. (2016). Can a Free Wearable Activity TrackerChange Behavior? The Impact of Trackers on Adults in a Physician-Led Wellness Group.*JMIR Research Protocols*, 5(4), e237. doi:10.2196/resprot.6534

- Gudzune, K., Doshi, R., Mehta, A., Chaudhry, Z., Jacobs, D., Vakil, R., ... Clark, J. (2015).
   Efficacy of commercial weight loss programs: an updated systematic review. *Annals of Internal Medicine*, *162*(7), 501–512. doi:10.7326/M14-2238
- Hruby, A., & Hu, F. (2015). The Epidemiology of Obesity: A Big Picture. *PharmacoEconomics*, *33*(7), 673–689. doi:10.1007/s40273-014-0243-x
- Kim, G., Lin, J., Blomain, E., & Waldman, S. (2014). Anti-Obesity Pharmacotherapy: New Drugs and Emerging Targets. *Clinical Pharmacology and Therapeutics*, 95(1), 53–66. doi:10.1038/clpt.2013.204
- Lewis, Z., Lyons, E., Jarvis, J., & Baillargeon, J. (2015). Using an electronic activity monitor system as an intervention modality: A systematic review. *BMC Public Health*, 15, 585. doi:10.1186/s12889-015-1947-3
- Maher, C., Ryan, J., Ambrosi, C., & Edney, S. (2017). Users' experiences of wearable activity trackers: a cross-sectional study. *BMC Public Health*, 17, 880. doi:10.1186/s12889-017-4888-1
- Naslund, J., Aschbrenner, K., Scherer, E., McHugo, G., Marsch, L., & Bartels, S. (2016).
  Wearable Devices and Mobile Technologies for Supporting Behavioral Weight Loss
  Among People with Serious Mental Illness. *Psychiatry Research*, 244, 139–144.
  doi:10.1016/j.psychres.2016.06.056

O'Brien, J., Gallagher, P., Stow, D., Hammerla, N., Ploetz, T., Firbank, M., . . . Olivier, P. (2017). A study of wrist-worn activity measurement as a potential real-world biomarker for late-life depression. *Psychological Medicine*, 47(1), 93-102. doi:10.1017/S0033291716002166

Office of Disease Prevention and Health Promotion. (2018). 2008 Physical Activity Guidelines

for Americans Summary. Retrieved from

https://health.gov/paguidelines/guidelines/summary.aspx

- Rosenberger, M., Buman, M., Haskell, W., McConnell, M., & Carstensen, L. (2016). 24 Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. *Medicine and Science in Sports and Exercise*, 48(3), 457–465.
  doi:10.1249/MSS.00000000000778
- Ross, K. & Wing, R. (2016). Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. *Obesity (Silver Spring)*, 24(8), 1653–1659. doi:10.1002/oby.21536
- Savage, N. (2017). The Measure of a Man. *Cell (Published by Elsevier Inc.), 169*, 1159-1161. doi:10.1016/j.cell.2017.05.037.
- Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H., Christle, J., Hastie, T., ... Ashley, E.
  (2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *Journal of Personalized Medicine*, 7(2), 3.
  doi:10.3390/jpm7020003
- Thompson, W., Kuhle, C., Koepp, G., McCrady-Spitzer, S., Levine, J. (2014). "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. *Archives of Gerontology and Geriatrics*, 58(3), 314-319. doi: 10.1016/j.archger.2014.01.004
- Townshend, T., & Lake, A. (2017). Obesogenic environments: current evidence of the built and food environments. *Perspectives in Public Health*, *137*(1), 38-44.
  doi:10.1177/1757913916679860
- Wallen, M., Gomersall, S., Keating, S., Wisloff, U., & Coombes, J. (2016). Accuracy of Heart

Rate Watches: Implications for Weight Management. *PLoS ONE*, *11*(5), e0154420. doi:10.1371/journal.pone.0154420

Wisconsin Department of Public Instruction. (2009). Wisconsin's Model Academic Standards for Nutrition. Retrieved from <u>https://dpi.wi.gov/sites/default/files/imce/team-</u>

nutrition/pdf/nestandards.pdf

Wisconsin Department of Public Instruction. (2010). Wisconsin Standards for Physical

Education. Retrieved from

https://dpi.wi.gov/sites/default/files/imce/sspw/pdf/pewiscstandards.pdf

- World Health Organization. (2018). Obesity and overweight. Retrieved from http://www.who.int/mediacentre/factsheets/fs311/en/
- World Health Organization. (2018). Physical Activity. Retrieved from

http://www.who.int/dietphysicalactivity/pa/en/

World Health Organization. (2018). Physical Inactivity: A Global Public Health Problem.

Retrieved from http://www.who.int/dietphysicalactivity/factsheet\_inactivity/en/

# APPENDIX A: QUALITY CRITERIA CHECKLISTS AND EVIDENCE ANALYSIS

# LIBRARY WORKSHEETS

# **Evidence Worksheet for Primary RESEARCH Article**

Citation: write it in AMA format	Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., & Actis, P.
as found in JADA.	(2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. <i>PLoS ONE</i> ,
	13(2), e0192691. doi:10.1371/journal.pone.0192691
Study design: Use algorithm –	Validity study
RCT, cohort, etc	
Study Class (A,B,C,D)	С
<b>Research Quality Rating</b> This rating tells if the research design is good $(+)$ , bad $(-)$ or neutral $(\emptyset)$ This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).	POSITIVE (+)
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the research question being investigated in the study?)	To assess in a controlled, research environment the accuracy and precision of the Fitbit Charge 2 for measuring heart rate (HR) with respect to a gold standard electrocardiograph
Inclusion criteria: requirements	Healthy adult participants
for study eligibility	
<b>Exclusion criteria</b> (conditions	Participants with neurological or cognitive disorders, recent musculoskeletal
<b>D</b> ecruitment	damage of surgery that would impair motor function, and fattoos
Recruitment Blinding used: some of the	NA
persons involved are prevented from knowing certain information that might lead to conscious or unconscious bias on their part, invalidating the results	
Description of study protocol	Participants rode a stationary bike for 10 minutes while their HR was
What happened in the study?	simultaneously recorded from each device
Intervention: Describe interventions, regimens, risk factors, or procedures studied.	For electrocardiograph recording, the electrode placement sites were prepared by standardized procedures of cleaning, shaving, and abrading the skin to improve signal acquisition and to minimize noise artifact. Three self-adhesive electrodes were placed on the upper torso. HR data per second was converted to bpm automatically by the data acquisition software program prior to analysis. The Fitbit Charge 2 was placed on the non-dominant wrist following manufacturer instructions and was charged fully prior to testing. Participants were asked to ride a stationary bike for 10 minutes with the stated goal to raise their HR as much as possible. Participants were free to slow down and rest at any time they desired to do so. The goal of the experiment was not to evaluate the training activity, but rather to collect enough HR data spanning as wide of a range of heart beats per minute (bpm) as possible. HR was acquired simultaneously using both devices (Fitbit Charge 2, ProComp Infiniti T7500M).

Statistical analysis: <i>List tests</i> .	Agreement between the Fitbit Charge 2 and the ProComp Infiniti T7500M was
significance level set a priori	estimated using the Bland-Altman method. The intraclass correlation coefficient
$(\alpha = 0.05; include intent to treat$	(ICC) was used as an alternative measure of agreement.
analysis if applicable; note if	
there is Power analysis.	
Timing of measurements: when	HR was measured continuously using both devices (Fitbit Charge 2, ProComp
outcomes were measured; usually	Infiniti T7500M) throughout the 10-minute intervention
baseline and one or more later	
times	
Dependent variables: outcomes	HR according to Fitbit Charge 2
that are measured or registered;	
variable whose change or	
different states the researcher	
wants to understand, explain, or	
predict	
Independent variables	Participants' level or intensity of cycling, instability or improper positioning of the
(intervention or procedure; this	devices
variable can be manipulated; a	
variable whose effect upon the	
dependent variable one is trying to	
understand)	
Control Variables	HR according to ProComp Infiniti T7500M
Examples: 1) multivariate logistic	
regression controlled for age,	
BMI, albumin;	
2) usual care; 3) isocaloric diet,	
etc.	
Initial n (e.g. 731 (298 males,	15 participants, 7 males and 8 females
433 females))	
Record number that entered study	
– not the number screened.	
Final n (attrition)	15 participants, 7 males and 8 females
number of subjects that completed	
	25.4.26
Age usually mean or range	25 to 36 years
Ethnicity (if given)	Lucasian
Other relevant demographics:	Unclear
aemographics describe the	
Anthronometrics: 2 2 2 200	Weight: 56 to 92 kg
Antin opometrics: e.g. were	Weight: 30 to 02 kg Height: 155 to 185 cm
groups sume or angerent on important physical measures	RMI: 20 to 25 kg/m <sup>2</sup>
(BML fitness level)	DWII. 20 to 25 kg/lii2
(Divil, funess level)	TSW VD Lab in Travico Italy
take place? City or country	15 w AF Lao III Heviso, Italy
Summary of Results: Abstract	The Fithit Charge 2 exhibited a mean hias of 5.9 hpm (95% CI: -6.1 to -5.6 hpm)
results including <i>quantitative data</i>	The limits of agreement $(I \circ A)$ which indicate the precision of individual
and statistics Include statistical	measurements, between the Fithit Charge 2 and ProComp Infiniti T7500M were
significance: P-values confidence	wide The upper LoA was +16.8 hpm whereas the lower LoA was -28.5 hpm
intervals (CI) relative risk (RR)	The ICC between the Fithit Charge 2 and ProComp Infiniti T7500M was 0.21
odds ratios (OR) likelihood ratio	$(95\% \text{ CI} \cdot 0.09 \text{ to } 0.34)$
number needed to treat nower	
analysis if available	
	Author's Conclusions

Author conclusion: paraphrase that stated by study author in body of the report or abstract	Findings are in line with those of several recent publications; the Fitbit Charge 2 presents an unchanged level of HR measurement accuracy compared to existing models of the same brand (i.e. Fitbit Charge HR). The Fitbit Charge 2 tends to underestimate HR, with moderate bias on average, although precision is poor for individual measurements, which could be underestimated by as much as 30 bpm. It may be speculated that the current algorithms for HR estimation lack proper sophistication.
Reviewer comments: Note strengths and limitations of study; identify concerns that affect study validity and generalizability— your comments should be italicized)	Strengths: zero percent attrition, stable positioning of the ProComp Infiniti T7500M Limitations: small sample size, participant recruitment was not discussed, blinding was not utilized, unstable positioning of the Fitbit Charge 2, lacking a defined activity pattern for the participants in order to simulate low, medium, and intensive exercise, and lacking a variety of participants with different skin tones, BMI, and ages Funding source: this work was supported by TSW XP Lab, which only provided financial support in the form of authors' salaries and/or research materials. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE QUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet) Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. <i>PLoS ONE</i> , <i>13</i> (2), e0192691.		Y E S	N O	U N C L E	N A
doi:10.1371/journal.pone.0192691				R	
1. Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some epidemiological studies)	1			X	
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/ population group would care about?	2	Х			
3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3		Х		
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "yes", the report is eligible for design the Evidence Quality Worksheet, depending on answers to the following validity questions.	ation w	ith a	plu	us (+	) on
VALIDITY QUESTIONS					
1. Was the <u>research question</u> clearly stated? This is usually stated at end of the introduction and just before methods section.		1	r 2 3	N 0 1	J N N A C L E A
		2	Κ	1	R
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified? <i>This is often called the treatment and explained in the methods section.</i>	1.1	2	K		
<ul> <li>1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?</li> <li>These are sometimes called the endpoints; the results section reports the outcomes, but this information should be in the methods section, too</li> </ul>	1.2	2	K		
1.3 Were the target population and setting specified? <i>The target population is group for whom findings may be applicable; look for this in the</i> <i>introduction and in the methods section</i>	1.3	2	K		
2. Was the <u>selection</u> of study subjects/patients free from bias?		1	2 3 7	N     O     	J N N A C L E A R
<ul> <li>2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?</li> <li>The authors should give several points about the inclusion/exclusion criteria such as the age range of the subjects, disease condition (like hyperlipidemia) required for inclusion. Exclusion criteria should be listed, too, although some are understood. For example if the ages for inclusion are 18 to 70, the authors will probably not specifically note that children and people over age 70 were excluded. Most of the time, however, subjects may be excluded for certain characteristics such as being pregnant or having some disease (like CHD).</li> <li>2.2 Were criteria applied equally to all study groups?</li> <li>2.3 Were health, demographics, and other characteristics of subjects described?</li> </ul>	2.1 2.2 2.2 2.3	2	X		
There is usually a Table 1 summarizing demographics and characteristics at baseline. Groups are <u>not</u> different if the P-Value is > 0.05. If there has been a previous paper describing the study population, that paper may be referenced and you would need to go back to the original publication to see that Table 1.	2.5		-		
2.4 Were the subjects/patients in a representative sample of the relevant population? The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion criteria.	2.4				X

3. \	Were <u>study groups comparable</u> ?		Y	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		s	Ŭ	C	
	Groups are <u>not</u> different if the P-Value is $> 0.05$ .				L E	
					A R	v
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				
	(Method of randomization identified if RCT)	5.1				Λ
	In a strong study, the authors may tell how the subjects were assigned to a group ( $e_{\alpha}$					
	randomized block design: or assigned by computer-generated random numbers)					
	Look for instances that show bias: for example I once read a study where patients					
	were randomized to receive liquid energy supplements: however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
	demographics) similar across study groups at baseline? See Table I for this - there					
	should be no significant differences across study groups in an intervention study.					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Х			
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
	prior study as a control group: that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control.					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4				Х
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
not applicable. Criterion may not be applicable in some cross-sectional studies.						
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	Х			
	reference standard (e.g. "gold standard")?					
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4.	Was method of handling <u>withdrawals</u> described?		Y E	N O	U N	N A
			s	-	C	
					E	
					A R	Х
	4.1 Were follow up methods described and the same for all groups?	4.1				Х
	4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up,	4.2				Х
	attrition rate) and/or response rate (cross-sectional studies) described for each group?					
	(Follow up goal for a strong study is 80 %.)					
	This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
	note that on the worksheet (as a note in the results section or in the reviewer					
	comments at the very bottom)					
	4.3 Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Х			
1	This information is often presented in a figure with # recruited, # enrolled (this is initial					
1	N), # remaining at end of study period (final N). Sometimes the reasons that subjects					
	withdrew or were dropped is given in the figure or in the text (results section).					
	4.4 Were reasons for withdrawals similar across groups?	4.4				Х
1		1	1	1		

	If there is a large attrition from one group and not others, you would want to look for a reason why; the answer to this question would then be no.					
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	X			
	The test under study? The test under study should be compared to reference test all the time. An example of this					
	might be using a DFX4 machine to measure percent body fat only if a subject's					
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30%$					
5	Was blinding used to prevent introduction of bias?		Y	N	U	N
			E S	0	N C L A R X	Α
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded to treatment group, <b>as appropriate</b> ?	5.1				Х
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MNT and who was not. Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	<ul> <li>5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).</li> </ul>	5.2			Х	
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	5.3				Х
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).					
	5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4				Х
	Establish who is a case and who is a control at the beginning of the study.					
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5			Х	
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?		Y E S	N O	U N C L E	N A
			Х		R	
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1				Х
	6.2 In observational study, were interventions, study settings, and clinicians/provider described?	6.2	Х			
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
	produce a meaningful effect?					
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)					
	6.4 Was the amount of exposure and if relevant subject/nations compliance measured?	64	x			
	How long did the treatment last? Did the patient follow directions?	0				
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5				Х
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)					
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall.					
	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7				Х

	For a study to be valid and unbiased, it is important that this be yes.					
	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8	X			
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E A	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X		R	
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				X
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul><li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li><li>Check that surveys were validated.</li></ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? Precision is reproducibility or repeatability	7.5	X			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S X	N O	U N C L E A R	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	X			
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals ( <i>mean</i> $\pm$ <i>CI</i> )	8.3	X			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4				X
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5				X

8.6 Was clinical significance as well as statistical significance reported? <i>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was</i> <i>reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical</i> <i>significance (P-value) and clinical significance (compare to standard of &lt; 200</i> <i>mg/do for normal cholesterol). A problem can occur when only statistical</i> <i>significance is reported. Reducing cholesterol from 300 to 250 might be statistically</i> <i>significant, but clinically it is still abnormal.</i>	8.6	X			
8.7 If negative findings, was a power calculation reported to address type 2 error? Type II ( $\beta$ error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."	8.7				Х
9. Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S	N O	U N C L	N A
		Х		E A R	
9.1 Is there a discussion of findings? Answer yes or no.	9.1	Х			
9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	Х			
<b>10.</b> Is bias due to study's <u>funding or sponsorship</u> unlikely? Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	Х			
<ul><li>10.2 Was there no apparent conflict of interest?</li><li>If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.</li></ul>	10.2	X			
SYMBOL					
MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report should be minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (0)	be design	ated	with	ı a	
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is excent	ionally st	rona	the	,	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

# PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

right. Academy of Nutrition and Dietetics

Citation	Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS ONE, 13(2), e0192691. doi:10.1371/journal.pone.0192691
Study Design	Validity study
Class	С
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\bigcirc$ (Neutral)
Research Purpose	To assess in a controlled research environment the accuracy and precision of the Fitbit Charge 2 for measuring heart rate (HR) with respect to a gold standard electrocardiograph
Inclusion Criteria	Healthy adult participants
Exclusion Criteria	Participants with neurological or cognitive disorders, recent musculoskeletal damage or surgery that would impair motor function, and tattoos
	Recruitment: Unclear
	Design: Participants rode a stationary bike for 10 minutes while their HR
	was simultaneously recorded from each device
	Blinding used (if applicable): NA
	Intervention (if applicable): For electrocardiograph recording, the
	electrode placement sites were prepared by standardized procedures of
	cleaning, shaving, and abrading the skin to improve signal acquisition and
	to minimize noise artifact. Three self-adhesive electrodes were placed on
Description of	the upper torso. HR data per second was converted to bpm automatically
Study Protocol	by the data acquisition software program prior to analysis. The Fitbit
	Charge 2 was placed on the non-dominant wrist following manufacturer
	instructions and was charged fully prior to testing. Participants were
	asked to ride a stationary bike for 10 minutes with the stated goal to raise
	their HR as much as possible. Participants were free to slow down and
	rest at any time they desired to do so. The goal of the experiment was not
	to evaluate the training activity, but rather to collect enough HR data
	spanning as wide of a range of heart beats per minute (bpm) as possible.

	HR was acquired simultaneously using both devices (Fitbit Charge 2,
	ProComp Infiniti T7500M).
	Statistical Analysis: Agreement between the Fitbit Charge 2 and the
	ProComp Infiniti T7500M was estimated using the Bland-Altman method.
	The intraclass correlation coefficient (ICC) was used as an alternative
	measure of agreement.
	Timing of Measurements: HR was measured continuously using both
	devices (Fitbit Charge 2, ProComp Infiniti T7500M) throughout the 10
Data Collection	minute intervention
Summary	Dependent Variables: HR according to Fitbit Charge 2
	Independent Variables: Participants' level or intensity of cycling,
	instability or improper positioning of the devices
	Control Variables: HR according to ProComp Infiniti T7500M
	Initial: 15 (7 Males 8 Females)
	Attrition (final N): 15
	Age: 25 to 36 years
Description of	Ethnicity: Caucasian
Actual Data	Other relevant demographics: Unclear
Sumple	Anthropometrics: Weight: 56 to 82 kg, Height: 155 to 185 cm, and BMI:
	20 to 25 kg/m2
	Location: TSW XP Lab in Treviso, Italy
	Key Findings: The Fitbit Charge 2 exhibited a mean bias of -5.9 bpm
	(95% CI: -6.1 to -5.6 bpm). The limits of agreement (LoA), which
	indicate the precision of individual measurements, between the Fitbit
	Charge 2 and ProComp Infiniti T7500M were wide. The upper LoA was
Summary of	+16.8 bpm, whereas the lower LoA was -28.5 bpm. The ICC between the
Kesults	Fitbit Charge 2 and ProComp Infiniti T7500M was 0.21 (95% CI: 0.09 to
	0.34).
	Other Findings:

	Findings are in line with those of several recent publications; the Fitbit
Author	Charge 2 presents an unchanged level of HR measurement accuracy
	compared to existing models of the same brand (i.e. Fitbit Charge HR).
	The Fitbit Charge 2 tends to underestimate HR, with moderate bias on
Conclusion	average, although precision is poor for individual measurements, which
	could be underestimated by as much as 30 bpm. It may be speculated that
	the current algorithms for HR estimation lack proper sophistication.
	Strengths: zero percent attrition, stable positioning of the ProComp
	Infiniti T7500M
	Limitations: small sample size, participant recruitment was not discussed,
Reviewer	blinding was not utilized, unstable positioning of the Fitbit Charge 2,
Comments	lacking a defined activity pattern for the participants in order to simulate
	low, medium, and intensive exercise, and lacking a variety of participants
	with different skin tones, BMI, and ages
	This work was supported by TSW XP Lab, which only provided financial
Funding Source	support in the form of authors' salaries and/or research materials. The
	funders had no role in study design, data collection and analysis, decision
	to publish, or preparation of the manuscript.

# Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
	<b>Positive</b> – Indicates that the report has clearly addressed issues of
т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	<i>Negative</i> – <i>Indicates that these issues have not been adequately addressed.</i>
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
U U	exceptionally week

Select a rating from the drop-down menu  $\psi$ 

Relevance Questions				
	1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear
	2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
	3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No
	4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Vê	Validity Questions					
1.	Was the <u>research question</u> clearly stated?	1	Yes			
	1.1. Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes			
	identified ? 1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes			
	1.3. Were the target population and setting specified?	1.3	Yes			
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Ves			
	2.1. Were inclusion/exclusion criteria specified (e.g., risk, point in disease	-	Veg			
	progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?	2.1	Yes			
	2.2. Were criteria applied equally to all study groups?	2.2	N/A			
	2.3. Were health, demographics, and other characteristics of subjects described?	2.3	Yes			
	2.4. Were the subjects/patients a representative sample of the relevant	2.4	Unalaar			
_	population?	2.4	Unclear			
3.	Were <u>study groups comparable</u> ? 3.1. Was the method of assigning subjects/patients to groups described and	3	N/A			
	unbiased? (Method of randomization identified if RCT)					
	3.2. Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.1	N/A			
	demographics) similar across study groups at baseline?					
	controls.)	3.2	N/A			
	3.4. If cohort study or cross-sectional study, were groups comparable on importan	33	Ves			
	confounding factors and/or were preexisting differences accounted for by using	5.5	105			
	3.5. If case control study, were potential confounding factors comparable for cases	3.4	N/A			
	and controls? (If case series or trial with subjects serving as own control, this					
	criterion is not applicable. Criterion may not be applicable in some cross-	3.5	N/A			
	3.6. If diagnostic test, was there an independent blind comparison with an					
	appropriate reference standard (e.g., "gold standard")?	3.6	Yes			
4.	Was method of handling withdrawals described?		N/A			
	4.1. Were follow up methods described and the same for all groups?	4 1	N/A			
	<ol> <li>was the number, characteristics of withdrawais (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> </ol>					
		4.2	N/A			
	4.3. Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Yes			
	4.4. Were reasons for withdrawals similar across groups	4.4	N/A			
	results of test under study?	4.5	Yes			
5.	Was blinding used to prevent introduction of bias?	_	- · · ·			
	5.1. In intervention study, were subjects, clinicians/practitioners, and investigators	5	Unclear			
	blinded to treatment group, as appropriate?	5.1	N/A			
	5.2. were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed					
	to be met.)	5.2	Unclear			
	5.3. In cohort study or cross-sectional study, were measurements of outcomes and	53	N/A			
1	risk factors blinded?	5.5	1N/ <i>F</i> <b>A</b>			

	5.4. In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4	N/A
	5.5. In diagnostic study, were test results blinded to patient history and other test results?	5.5	Unclear
6.	Were <u>intervention</u> /therapeutic regimens/exposure factor or procedure and any	6	Yes
	comparison(s) described in detail? Were <u>intervening factors</u> described?		
	6.1. In RCI or other intervention trial, were protocols described for all regimens studied?	6.1	N/A
	6.2. In observational study, were interventions, study settings, and	6.2	Yes
	6.3. Was the intensity and duration of the intervention or exposure factor sufficient	6.3	Yes
	to produce a meaningful effect?	6.4	Yes
	measured?	6.5	N/A
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.6	N/A
	6.7. Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?	6.7	N/A
	6.8. In diagnostic study, were details of test administration and replication	6.8	Vas
7	Were outcomes clearly defined and the measurements valid and reliable?	0.0	Tes Vac
<i>.</i>	7.1. Were primary and secondary endpoints described and relevant to the	7 1	Yes
	question?	7.1	
	7.3 Was the period of follow-up long enough for important outcome(s) to occur?	1.2	
	7.4. Were the observations and measurements based on standard valid and	7.3	N/A
	reliable data collection instruments/tests/procedures?	7.4	Yes
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	Yes
		7.7	N/A
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome	8	Yes
	8.1. Were statistical analyses adequately described the results reported	8.1	Yes
	appropriately? 8.2 Were correct statistical tests used and assumptions of test not violated?	8.2	Yes
	8.3. Were statistics reported with levels of significance and/or confidence intervals?	83	Ves
	8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response	0.5 0.4	
	analysis)?	0.4	
	8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?	8.5	N/A
	8.6. Was clinical significance as well as statistical significance reported?	8.6	Yes
	8.7. If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A
9.	Are <u>conclusions supported by results</u> with biases and limitations taken into	9	Yes
	consideration?	9.1	Yes
	9.1. Is there a discussion of indings? 9.2. Are biases and study limitations identified and discussed?	9.2	Yes
10.	Is bias due to study's <u>funding or sponsorship</u> unlikely?	10	Yes
	10.1. Were sources of funding and investigators' affiliations described?	10.1	Yes
	10.2. Was there no apparent conflict of interest?	10.2	Yes
MI	NUS/NEGATIVE (-)		

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

# NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

# PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

Citation write it in AMA format as	Codmus Portrom I. Congnon P. Wirkus F. I. Throon Porowski V. M. &				
Citation: write it in AMA format as	Caunius-Deluani, L., Gangnoni, K., Wirkus, E. J., Thiath-Dolowski, K. M., &				
Jouna in JADA.	Gorzentz-Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by				
	Some wrist-worn Activity I rackers. Annals of Internal Medicine, 166(8), 610–				
	012. d01:10./320/L10-0353				
Study design: Use algorithm –	Validity study				
RCT, cohort, etc	~				
Study Class (A,B,C,D)	C				
<b>Research Quality Rating</b>	NEUTRAL (Ø)				
This rating tells if the research					
design is good (+), bad (-) or					
neutral ( $\varnothing$ )					
This is determined by the quality					
criteria list. Delete the ratings that					
do not apply (i.e. if positive, delete					
minus/negative and neutral).					
· · · · · · · · · · · · · · · · · · ·	Purpose/Population Studied/Practice Studied				
<b>Research purpose:</b> What is the	To determine the accuracy of the heart rate measured by four commercial light-				
research auestion being	emitting diode–dependent, wrist-worn activity trackers				
investigated in the study?)	ennung aloue aepenaent, wist worn aeuwig auereis				
Inclusion criteria: naquinomenta for	Healthy adult participants				
study eligibility	Treating adult participants				
<b>Exclusion criteria</b> (conditions that	Participants with cardiovascular conditions				
make individual ineligible)	i articipants with cardiovascular conditions				
Recruitment	Unclear				
Rect utilient Rlinding used: some of the persons	NΔ				
involved are prevented from	1177				
knowing cartain information that					
might lead to conscious or					
unconscious bias on their part					
invalidating the results					
Description of study protocol	Four activity trackers (Basis Back Fithit Charge Fithit Surge Mie Fuse) were				
What happened in the study?	selected and tested for resting and active heart rate measurement accuracy				
what happened in the study?	science and rested for resting and active heart face measurement accuracy				
Interventions Describe	Ear and participant two activity trackers were placed on each wrist. Next				
intervention: Describe	For each participant, two activity trackers were placed on each wrist. Next,				
interventions, regimens, risk jaciors,	has the sector in the sector of the sector o				
or procedures studied.	nearl rates were measured at one-minute intervals for 10 minutes. Then,				
	participants heart rates were measured at one-minute intervals for 10 minutes				
Statistical analysis I've	Nine they exercised on a treadmin at 05 percent of their maximum neart rate.				
Statistical analysis: List tests,	biand-Animan piors were used to compare the neart rates measured by the				
significance level set a priori	electrocardiograph and by each of the activity trackers				
$(\alpha=0.05;$ include intent to treat					
analysis if applicable; note if there					
is Power analysis.					
1 iming of measurements: when	Heart rate was measured at one-minute intervals throughout the 20-minute				
ourcomes were measured; usually	Fick it Change Fick it Summer Ministry trackers (Basis Peak,				
baseline and one or more later	riton Charge, Fitolt Surge, Milo Fuse)				
times					
Dependent variables: outcomes	Heart rate according to Basis Peak, Fitbit Charge, Fitbit Surge, Mio Fuse				
that are measured or registered;					
variable whose change or different					
states the researcher wants to					
understand, explain, or predict					

Independent variables	Participants' level of physical fitness
(intervention or procedure; this	
variable can be manipulated; a	
variable whose effect upon the	
dependent variable one is trying to	
understand)	
Control Variables	Heart rate according to electrocardiograph
Examples: 1) multivariate logistic	
regression controlled for age,	
BMI, albumin;	
2) usual care; 3) isocaloric diet,	
etc.	
<b>Initial n</b> (e.g. 731 (298 males,	40 participants, 20 males and 20 females
433 females))	
Record number that entered study	
<i>– not the number screened.</i>	
<b>Final n</b> (attrition)	40 participants, 20 males and 20 females
number of subjects that completed	
study	
Age usually mean or range	30 to 65 years
Ethnicity (if given)	Unclear
Other relevant demographics:	Unclear
demographics describe the	
population (students, athletes, etc)	
Anthropometrics: e.g. were	Mean BMI: 25.1 kg/m2
groups same or all gerent on	
(PML fitness level)	
(DMI, filless level)	Unaloar
take place? City or country	Olicical
Summary of Posults: Abstract	For participants at ract, the limits of agreement was best for the Fithit Surge, which
results including <i>quantitative data</i>	had the narrowest limits of agreement $(-5.1 \text{ to } 4.5 \text{ heats/min})$ worst for the Basis
and statistics. Include statistical	Part (-17, 1, to, 22, 6, heats/min) and intermediate for the Eithit Charge (-10, 5, to, 0, 2)
significance: P-values confidence	heats/min) and Mio Fuse $(-7.8 \text{ to } 9.9 \text{ heats/min})$ . When participants exercised at
intervals (CI) relative risk (RR)	65% of their maximum heart rate the limits of agreement were relatively noor for
odds ratios (OR) likelihood ratio	all the activity trackers (Mio Fuse $-22.5$ to 26.0 heats/min: Basis Peak $-27.1$ to
number needed to treat, power	29.2 beats/min <sup>-</sup> Fitbit Surge -34.8 to 39.0 beats/min <sup>-</sup> and Fitbit Charge -41.0 to
analysis if available.	36.0 beats/min). The repeatability coefficient for the electrocardiograph was 5.3
	beats/min at rest and 9.1 beats/min during exercise. In comparison, the
	repeatability coefficient at rest was 4.2 beats/min for the Fitbit Surge, 9.3 beats/min
	for the Fitbit Charge, 10.9 beats/min for the Mio Fuse, and 19.3 beats/min for the
	Basis Peak. During exercise, the repeatability coefficient was 20.2 beats/min for
	the Basis Peak, 20.6 beats/min for the Fitbit Surge, 21.6 beats/min for the Fitbit
	Charge, and 23.7 beats/min for the Mio Fuse.
	Author's Conclusions
Author conclusion: paraphrase	Some of the activity trackers measured values for heart rate that were similar to
that stated by study author in	those measured by the electrocardiograph. All of the activity trackers were more
body of the report or abstract	accurate at rest than during moderate exercise, performance at rest was better for
	some trackers than for others, and limited repeatability for each tracker caused
	more problems than poor agreement between each activity tracker and the
	electrocardiograph. Although current trackers may help persons self-monitor their
	daily activity, more research is needed before we can confidently conclude that the
	monitoring feature for heart rate is sufficient to help clinicians advise their patients

	about health issues and conduct clinical trials that require a high level of accuracy			
	and reliability for heart rate measurement.			
Reviewer comments: Note	Strengths: zero percent attrition, defined activity pattern for participants to			
strengths and limitations of study;	stimulate minimal and moderate exercise			
identify concerns that affect study	Limitations: participant recruitment, demographics, and location of study were not			
validity and generalizability—	discussed, and blinding was not utilized			
your comments should be	Funding source: unclear, authors disclosed no conflicts of interest			
italicized)				

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE QUESTIONS						
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)		Y E	N O	U N	N A	
Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-		s		C		
Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity				E		
Trackers. Annals of Internal Medicine, 166(8), 610–612. doi:10.7326/L16-0353				A R		
1. Would implementing the studied intervention or procedure (if found successful) result in	1			Х		
improved outcomes for the patients/clients/population group? (Not Applicable for some						
epidemiological studies)						
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	Х				
population group would care about?						
3. Is the focus of the intervention or procedure (independent variable) or topic of study a	3		Х			
common issue of concern to dietetics practice?						
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х				
If the answers to all of the above relevance questions are "yes", the report is eligible for design	ation w	vith a	plus	5 (+)	on	
the Evidence Quality Worksheet, depending on answers to the following validity questions.						
VALIDITY QUESTIONS						
1. Was the <u>research question</u> clearly stated?		Y	N C	N U	N	
This is usually stated at end of the introduction and just before methods section.		s				
				E		
		Х		AR		
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified?	11	x				
This is often called the treatment and explained in the methods section		11				
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?	12	x				
These are sometimes called the endpoints: the results section reports the outcomes, but	1.2	11				
this information should be in the methods section, too						
1.3 Were the target population and setting specified?	1.3	X				
The target population is group for whom findings may be applicable; look for this in the						
introduction and in the methods section						
2. Was the selection of study subjects/patients free from bias?		Y	ľ		N	
		S			A	
				L E		
				A		
				Σ	Κ	
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1	Х	-			
diagnostic or prognosis criteria), and with sufficient detail and without omitting						
criteria critical to the study?						
The authors should give several points about the inclusion/exclusion criteria such as the						
age range of the subjects, disease condition (like hyperlipidemia) required for						
inclusion. Exclusion criteria should be listed, too, although some are						
understood. For example if the ages for inclusion are 18 to 70, the authors will						
probably not specifically note that children and people over age 70 were						
excluded. Most of the time, however, subjects may be excluded for certain						
characteristics such as being pregnant or having some disease (like CHD).			_			
2.2 Were criteria applied equally to all study groups?	2.2				X	
2.3 Were health, demographics, and other characteristics of subjects described?	2.3			X		
There is usually a Table I summarizing demographics and characteristics at baseline.						
Groups are <u>not</u> afferent if the P-Value is $> 0.05$ . If there has been a previous						
paper aescribing the study population, that paper may be rejerenced and you						
would need to go back to the original publication to see that Table 1.	2.4		_		7	
2.4 were the subjects/patients in a representative sample of the relevant population?	2.4			1	`	
	The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion outoring					
-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	--------	----------	--------	---
3 1	Criteria. Wara study groups comparable?	+	Y	N	U	N
5.	There is usually a Table 1 summarizing demographics and characteristics at baseline		E	0	N C	А
	Groups are not different if the P-Value is > 0.05		5		L	
	Groups are <u>not</u> afferent if the 1 + affects > 0.00.				A R	Х
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!	2.2	_			V
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
	demographics) similar across study groups at baseline? See Table 1 for this - there					
	2.2 Were consumption of the sector of the se	2.2	v			
	S.5 were concurrent controls used? (Concurrent preferred over instorical controls.)	3.3	Λ			
	most RCTs use a control group: that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	34				X
	confounding factors and/or were preexisting differences accounted for by using	5.1				
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					L
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.	2.6				
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	X			
	reference standard (e.g. "gold standard")?					
	Example: comparing boay jai analysis method with underwater weigning (gold					
	Standard). In studies if ying to determine the best equation (like Mijjiin-St. Jeor of Hawis Repediet) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used					
4.	Was method of handling withdrawals described?	-	Y	N	U	N
	was memore of nanoming <u>withdrawars</u> described.		E S	0	N C	А
					L E	
					A	Х
	4.1 Were follow up methods described and the same for all groups?	4.1			ĸ	Х
	4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up,	4.2				Х
	attrition rate) and/or response rate (cross-sectional studies) described for each group?					
	(Follow up goal for a strong study is 80 %.)					
	This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
	note that on the worksheet (as a note in the results section or in the reviewer					
	<i>comments at the very bottom)</i>	4.2	37	<u> </u>		
	4.5 were all enrolled subjects/patients (in the original sample) accounted for?	4.5	X			
1	<i>I his information is often presented in a figure with # recruited, # enrolled (this is initial N), # remaining at and of study navied (final N). Sometimes the reasons that subjects</i>					
1	$T_{ij}$ , $\pi$ remaining at end of study period (final $T_{ij}$ ). Sometimes the reasons that subjects withdrew or were dropped is given in the figure or in the text (results section)					
	manufer of were an opped is given in the jighte of the the text (results section).	1		1		

	A.A. Were reasons for withdrawals similar across groups?	11				v
	4.4 Were reasons for withdrawars similar across groups: If there is a large attrition from one group and not others, you would want to look for a	4.4				Λ
	If there is a targe attrition from one group and not others, you would want to took for a					
	A 5 16 discreption text and desiries to perform references text and denter dent or results of	15	v			
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	Λ			
	test under study?					
	The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					
	BMI was $>$ 35 but bioimpedance analyzer indicated body fat $<$ 30%.					0
5.	Was <u>blinding</u> used to prevent introduction of bias?		Y E	N O	U N	A
			s		C	
					E	
					A	
					X	
	5.1 In intervention study were subjects clinicians/practitioners and investigators blinded	51	1			x
	to treatment group as appropriate?	5.1				21
	The key term is an appropriate. For example, in the Lim et al 2008 study, the investigators					
	The key term is as appropriate. For example, in the Lim et al 2006 study, the investigators					
	sudied the effect of MNT on tipla levels in hypercholesterotemic patients. It was					
	an RC1, but obviously, the subjects and practitioners knew who was getting MINT					
	and who was not. Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2			Х	
	using an objective test, such as a lab value, this criterion is assumed to be met.)					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					1
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer ves, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					1
	5.4 In case control study, was case definition explicit and case ascertainment not	54	1			X
	influenced by exposure status?	5.1				21
	Establish who is a case and who is a control at the beginning of the study					
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5			v	
	5.5 In diagnostic study, were test results officied to patient instory and other test results?	5.5	v	N		N
0.	were intervention/therapeutic regimens/exposure factor or procedure and any		Ē	0	Ň	A
	comparison(s) described in detail? Were intervening factors described?		s		C L	
					E	
			X		R	
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1				Х
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2	Х			
	described?					
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	63	x			
	nroduce a meaningful effect?	0.5	11			
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
	difference in lab values for cholesteral; however 12 days would not be long					
	anguah)					
	(A Was the amount of expressive and if relevant subject/notions compliance managured?	6.4	v			
	Use was the amount of exposure and, if relevant, subject/patient compliance measured?	0.4	Λ			
	How long and the treatment tast? Did the patient joitow directions?	6.5				37
	0.5 were co-interventions (e.g., ancillary treatments other therapies) described?	0.5	1			Х
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)		<u> </u>			<b>-</b>
	6.6 Were extra or unplanned treatments described?	6.6	1			Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the		1			i
	discussion section. It is likely there were no unplanned treatments, so a "no"		1			
	answer is not a problem overall.					

	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups? <i>For a study to be valid and unbiased, it is important that this be yes.</i>	6.7				Х
	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8	Х			
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E A	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X		R	
	7.2 Were nutrition measures appropriate to question and outcomes of concern? Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? Precision is reproducibility or repeatability	7.5	Х			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7			_	Χ
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S X	N O	U N C L E A R	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	X			
	<ul> <li>8.2 Were correct statistical tests used and assumptions of test not violated?</li> <li>You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).</li> </ul>	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals ( <i>mean</i> $\pm$ <i>CI</i> )	8.3			Х	
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the</li> </ul>	8.4				X
	statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.					

<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li><i>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical significance is reported. Reducing cholesterol from 300 to 250 might be statistically significant, but clinically it is still abnormal.</i></li> </ul>	8.6			X	
<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li>Type II (β error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				Х
9. Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S	N O	U N C L E	N A
		Х		A R	
9.1 Is there a discussion of findings? Answer yes or no.	9.1	Х			
9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2			Х	
<b>10.</b> Is bias due to study's <u>funding or sponsorship</u> unlikely? Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L A R	N A
<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1			Х	
<ul><li>10.2 Was there no apparent conflict of interest?</li><li>If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.</li></ul>	10.2	X			
SYMBOL					
MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report should minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (a)	be design	ated	with	h a	
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is excern	tionally si	trong	the	,	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Citation	Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. Annals of Internal Medicine, 166(8), 610–612. doi:10.7326/L16-0353
Study Design	Validity study
Class	С
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\boxtimes \otimes$ (Neutral)
Research Purpose	To determine the accuracy of the heart rate measured by four commercial, light-emitting diode–dependent, wrist-worn activity trackers
Inclusion Criteria	Healthy adult participants
Exclusion Criteria	Participants with cardiovascular conditions
	Recruitment: Unclear
	Design: Four activity trackers (Basis Peak, Fitbit Charge, Fitbit Surge,
	Mio Fuse) were selected and tested for resting and active heart rate
	measurement accuracy against a reference standard electrocardiograph
	Blinding used (if applicable): NA
	Intervention (if applicable): For each participant, two activity trackers
Description of	were placed on each wrist. Next, participants were seated, connected to
Study Protocol	an electrocardiograph, and their resting heart rates were measured at one-
	minute intervals for 10 minutes. Then, participants' heart rates were
	measured at one-minute intervals for 10 minutes while they exercised on a
	treadmill at 65 percent of their maximum heart rate.
	Statistical Analysis: Bland-Altman plots were used to compare the heart
	rates measured by the electrocardiograph and by each of the activity
	trackers
	Timing of Measurements: Heart rate was measured at one-minute
Data Collection	intervals throughout the 20-minute intervention using the
Summary	electrocardiograph and four activity trackers (Basis Peak, Fitbit Charge,
	Fitbit Surge Mio Fuse)
	1 100 Sulpo, 1110 1 100)

	Dependent Variables: Heart rate according to Basis Peak, Fitbit Charge,
	Fitbit Surge, Mio Fuse
	Independent Variables: Participants' level of physical fitness
	Control Variables: Heart rate according to electrocardiograph
	Initial: 40 (20 Males 20 Females)
	Attrition (final N): 40
Description of	Age: 30 to 65 years
Actual Data	Ethnicity: Unclear
Sample	Other relevant demographics: Unclear
	Anthropometrics: Mean BMI: 25.1 kg/m2
	Location: Unclear
	Key Findings: For participants at rest, the limits of agreement was best for
	the Fitbit Surge, which had the narrowest limits of agreement $(-5.1 \text{ to } 4.5 \text{ to } $
	beats/min), worst for the Basis Peak (-17.1 to 22.6 beats/min), and
	intermediate for the Fitbit Charge (-10.5 to 9.2 beats/min) and Mio Fuse
	(-7.8 to 9.9 beats/min). When participants exercised at 65% of their
	maximum heart rate, the limits of agreement were relatively poor for all
	the activity trackers (Mio Fuse, -22.5 to 26.0 beats/min; Basis Peak,
	-27.1 to 29.2 beats/min; Fitbit Surge, -34.8 to 39.0 beats/min; and Fitbit
Summary of	Charge, -41.0 to 36.0 beats/min). The repeatability coefficient for the
Results	electrocardiograph was 5.3 beats/min at rest and 9.1 beats/min during
	exercise. In comparison, the repeatability coefficient at rest was 4.2
	beats/min for the Fitbit Surge, 9.3 beats/min for the Fitbit Charge, 10.9
	beats/min for the Mio Fuse, and 19.3 beats/min for the Basis Peak.
	During exercise, the repeatability coefficient was 20.2 beats/min for the
	Basis Peak, 20.6 beats/min for the Fitbit Surge, 21.6 beats/min for the
	Fitbit Charge, and 23.7 beats/min for the Mio Fuse.
	Other Findings:
Author	Some of the activity trackers measured values for heart rate that were
Conclusion	similar to those measured by the electrocardiograph. All of the activity

	trackers were more accurate at rest than during moderate exercise,
	performance at rest was better for some trackers than for others, and
	limited repeatability for each tracker caused more problems than poor
	agreement between each activity tracker and the electrocardiograph.
	Although current trackers may help persons self-monitor their daily
	activity, more research is needed before we can confidently conclude that
	the monitoring feature for heart rate is sufficient to help clinicians advise
	their patients about health issues and conduct clinical trials that require a
	high level of accuracy and reliability for heart rate measurement.
	Strengths: zero percent attrition, defined activity pattern for participants
Reviewer	to stimulate minimal and moderate exercise
Comments	Limitations: participant recruitment, demographics, and location of study
	were not discussed, and blinding was not utilized
Funding Source	Unclear, authors disclosed no conflicts of interest

### Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
0	exceptionally week

Select a rating from the drop-down menu ↓

Re	elevance Questions		
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Validity Questions		
1. Was the <u>research question</u> clearly stated?	1	Yes

		1.1.	Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes
		1 2	identified ? Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
		1.2.	Were the target population and setting specified?	1.3	Yes
Ē	2.	Was t	ne <u>selection</u> of study subjects/patients free from bias?	2	Unclear
		2.1.	Were inclusion/exclusion criteria specified (e.g., risk, point in disease	2.1	Vaa
			progression, diagnostic or prognosis criteria), and with sufficient detail and witheut emitting criteria critical to the study?	2.1	res
		2.2.	Were criteria applied equally to all study groups?	2.2	N/A
		2.3.	Were health, demographics, and other characteristics of subjects described?	2.3	No
		2.4.	Were the subjects/patients a representative sample of the relevant	2.4	Unalaar
				2.4	Ulicitai
	3.	Were	study groups comparable?	3	N/A
		3.1.	was the method of assigning subjects/patients to groups described and unbiased? (Method of randomization identified if RCT)		
		3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.1	N/A
			demographics) similar across study groups at baseline?		
		3.3.	Were concurrent controls used? (Concurrent preferred over historical	3.2	N/A
		2.4	controls.) If cohort study or cross sectional study, were groups comparable on important		
		5.4.	confounding factors and/or were preexisting differences accounted for by using	3.3	Yes
			appropriate adjustments in statistical analysis?		
		3.5.	If case control study, were potential confounding factors comparable for cases	3.4	N/A
			and controls? (If case series or trial with subjects serving as own control, this		
			sectional studies )	3.5	N/A
		3.6.	If diagnostic test, was there an independent blind comparison with an	2.6	N/
			appropriate reference standard (e.g., "gold standard")?	3.6	Yes
- 1		Was n	nethod of handling withdrawals described?		N/A
	4.			4	1 1/ / 1
	4.	4.1.	Were follow up methods described and the same for all groups?	4	N/A
	4.	4.1. 4.2.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for	4	N/A
	4.	4.1. 4.2.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)	4 4.1 4.2	N/A N/A
	4.	4.1. 4.2. 4.3.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for?	4 4.1 4.2 4.3	N/A N/A Yes
	4.	4.1. 4.2. 4.3. 4.4.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups	4 4.1 4.2 4.3 4.4	N/A N/A Yes N/A
	4.	4.1. 4.2. 4.3. 4.4. 4.5.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups If diagnostic test, was decision to perform reference test not dependent on recult of test under study?	4 4.1 4.2 4.3 4.4	N/A N/A Yes N/A
	4.	4.1. 4.2. 4.3. 4.4. 4.5.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4 4.1 4.2 4.3 4.4 4.5	N/A N/A Yes N/A Yes
	4. 5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was b</b> 5 1	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups If diagnostic test, was decision to perform reference test not dependent on results of test under study? <b>linding used to prevent introduction of bias?</b>	4 4.1 4.2 4.3 4.4 4.5 5	N/A N/A Yes N/A Yes Unclear
-	4. 5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups If diagnostic test, was decision to perform reference test not dependent on results of test under study? <b>linding used to prevent introduction of bias?</b> In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?	4 4.1 4.2 4.3 4.4 4.5 5	N/A N/A Yes N/A Yes Unclear
	4. 5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.) Were all enrolled subjects/patients (in the original sample) accounted for? Were reasons for withdrawals similar across groups If diagnostic test, was decision to perform reference test not dependent on results of test under study? <b>linding used to prevent introduction of bias?</b> In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is	4 4.1 4.2 4.3 4.4 4.5 5 5.1	N/A N/A Yes N/A Yes Unclear N/A
	4.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Iinding used to prevent introduction of bias?</li> <li>In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.1 5.2	N/A N/A Yes N/A Yes Unclear N/A
	4.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Inding used to prevent introduction of bias?</li> <li>In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2	N/A N/A Yes N/A Yes Unclear N/A Unclear
	4.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2. 5.3.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Inintervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2 5.3	N/A N/A Yes N/A Yes Unclear N/A Unclear N/A
	4.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2. 5.3. 5.4.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Inintervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2 5.3 5.3	N/A N/A Yes N/A Yes Unclear N/A Unclear N/A
	5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2. 5.3. 5.4.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Iinding used to prevent introduction of bias?</li> <li>In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li> <li>In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2 5.3 5.4	N/A N/A Yes N/A Yes Unclear N/A Unclear N/A N/A
	5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2. 5.3. 5.4. 5.5.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li> <li>In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>In diagnostic study, were test results blinded to patient history and other test results?</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2 5.3 5.4 5.5	N/A N/A Yes N/A Yes Unclear N/A Unclear N/A N/A Unclear
	5.	4.1. 4.2. 4.3. 4.4. 4.5. <b>Was <u>b</u></b> 5.1. 5.2. 5.3. 5.4. 5.5.	<ul> <li>Were follow up methods described and the same for all groups?</li> <li>Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80%.)</li> <li>Were all enrolled subjects/patients (in the original sample) accounted for?</li> <li>Were reasons for withdrawals similar across groups</li> <li>If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>Iinding used to prevent introduction of bias?</li> <li>In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate?</li> <li>Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li> <li>In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>In diagnostic study, were test results blinded to patient history and other test results?</li> </ul>	4 4.1 4.2 4.3 4.4 4.5 5 5.1 5.2 5.3 5.4 5.5	N/A N/A Yes N/A Yes Unclear N/A Unclear N/A N/A Unclear

6.	Were intervention/therapeutic regimens/exposure factor or procedure and any	6.1	N/A
	comparison(s) described in detail? Were <u>intervening factors</u> described?		
	6.1. In RCT or other intervention trial, were protocols described for all regimens studied?	6.2	Yes
	6.2. In observational study, were interventions, study settings, and	6.3	Yes
	clinicians/provider described? 6.3 Was the intensity and duration of the intervention or exposure factor sufficient	6.4	Yes
	to produce a meaningful effect?	6.5	N/A
	6.4. Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.6	N/A
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described? 6.6. Were extra or unplanned treatments described?	6.7	N/A
	6.7. Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?		
	6.8. In diagnostic study, were details of test administration and replication sufficient?	6.8	Yes
7.	Were outcomes clearly defined and the measurements valid and reliable?	7	Yes
	7.1. Were primary and secondary endpoints described and relevant to the	7 1	V
	question?	/.1	Yes
	7.2. Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A
	7.3. Was the period of follow-up long enough for important outcome(s) to occur? 7.4. Were the observations and measurements based on standard, valid, and	7.3	N/A
	reliable data collection instruments/tests/procedures?	7.4	Yes
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	Vec
	7.7. Were the measurements conducted consistently across groups?	7.7	N/A
8.	Was the statistical analysis appropriate for the study design and type of outcome	0	Vaa
	indicators?	0	res
	8.1. Were statistical analyses adequately described the results reported	8.1	Yes
	appropriately?		
1	appropriately? 8.2. Were correct statistical tests used and assumptions of test not violated?	8.2	Yes
	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate was there</li> </ul>	8.2 8.3	Yes Unclear
	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response</li> </ul>	8.2 8.3 8.4	Yes Unclear N/A
	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that</li> </ul>	8.2         8.3         8.4         8.5	Yes Unclear N/A N/A
	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> </ul>	8.2         8.3         8.4         8.5         8.6	Yes Unclear N/A N/A Unclear
	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> </ul>	8.2           8.3           8.4           8.5           8.6           8.7	Yes Unclear N/A N/A Unclear N/A
9.	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> </ul> Are conclusions supported by results with biases and limitations taken into	8.2         8.3         8.4         8.5         8.6         8.7         9	Yes Unclear N/A N/A Unclear N/A Yes
9.	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> </ul> Are <u>conclusions supported by results</u> with biases and limitations taken into consideration?	8.2 8.3 8.4 8.5 8.6 8.7 9 9	Yes Unclear N/A N/A Unclear N/A Yes Yes
9.	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> </ul> Are conclusions supported by results with biases and limitations taken into consideration? <ul> <li>9.1. Is there a discussion of findings?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1	Yes Unclear N/A N/A Unclear N/A Yes Yes
9.	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear
9. <b>10.</b>	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes
9. <b>10.</b>	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> <li>Is bias due to study's funding or sponsorship unlikely?</li> <li>10.1. Were sources of funding and investigators' affiliations described?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10         10.1	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes Unclear
9. <b>10.</b>	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> <li>Is bias due to study's funding or sponsorship unlikely?</li> <li>10.1. Were sources of funding and investigators' affiliations described?</li> <li>10.2. Was there no apparent conflict of interest?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10         10.1         10.2	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes Unclear Yes
9. 10. MI	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> <li>Is bias due to study's funding or sponsorship unlikely?</li> <li>10.1. Were sources of funding and investigators' affiliations described?</li> <li>NUS/NEGATIVE (-)</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10         10.1         10.2	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes Unclear Yes
9. 10. MI <i>If n</i>	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> <li>Is bias due to study's funding or sponsorship unlikely?</li> <li>10.1. Were sources of funding and investigators' affiliations described?</li> <li>10.2. Was there no apparent conflict of interest?</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10         10.1         10.2         Id be determined	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes Unclear Yes Unclear Yes
9. 10. MI If m (-)	<ul> <li>appropriately?</li> <li>8.2. Were correct statistical tests used and assumptions of test not violated?</li> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>8.5. Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?</li> <li>8.6. Was clinical significance as well as statistical significance reported?</li> <li>8.7. If negative findings, was a power calculation reported to address type 2 error?</li> <li>Are conclusions supported by results with biases and limitations taken into consideration?</li> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations identified and discussed?</li> <li>Is bias due to study's <u>funding or sponsorship</u> unlikely?</li> <li>10.1. Were sources of funding and investigators' affiliations described?</li> <li>10.2. Was there no apparent conflict of interest?</li> <li>NUS/NEGATIVE (-)</li> <li>tost (six or more) of the answers to the above validity questions are "No," the report show symbol on the Evidence Worksheet.</li> </ul>	8.2         8.3         8.4         8.5         8.6         8.7         9         9.1         9.2         10         10.1         10.2         Id be determined	Yes Unclear N/A N/A Unclear N/A Yes Yes Unclear Yes Unclear Yes Unclear Yes

NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Cadmus-Bertram, L., Marcus, B., Patterson, R., Parker, B., & Morey, B. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. <i>American Journal of Preventive Medicine</i> , <i>49</i> (3), 414–418. doi:10.1016/j.amepre.2015.01.020
<b>Study design:</b> Use algorithm – <i>RCT, cohort, etc</i>	Randomized controlled trial
Study Class (A,B,C,D)	А
<b>Research Quality Rating</b> This rating tells if the research design is good (+), bad (-) or neutral (Ø) This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).	POSITIVE (+)
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the research question being investigated in the study?)	To evaluate, within a randomized controlled trial, the feasibility and preliminary efficacy of integrating a Fitbit tracker and website into a physical activity intervention for postmenopausal, overweight/obese women
Inclusion criteria: requirements for study eligibility	Postmenopausal women, overweight/obese, perform less than 60 minutes per week of moderate-to-vigorous intensity physical activity (MVPA), could exercise safely, were regular internet users, and owned a tablet/computer
<b>Exclusion criteria</b> (conditions that make individual ineligible)	Women who responded, "somewhat uncomfortable," "very uncomfortable," "somewhat do not enjoy," or "very much do not enjoy" on baseline questionnaire
Recruitment	Unclear, supported by the Athena Breast Health Network
Blinding used: some of the persons involved are prevented from knowing certain information that might lead to conscious or unconscious bias on their part, invalidating the results	Unclear
<b>Description of study protocol</b> What happened in the study?	Participants were randomized to either a Fitbit or pedometer-based intervention group to determine whether the Fitbit One increased physical activity more than the pedometer in postmenopausal, overweight/obese women
Intervention: Describe interventions, regimens, risk factors, or procedures studied.	Participants received a baseline questionnaire, anthropometric measurements, and physical activity was measured for seven days using the ActiGraph GT3X+ accelerometer. Then, participants were randomized to a 16-week web-based self-monitoring intervention (N=25) or a comparison group (N=26). Participants in the Web-Based Tracking Group received a Fitbit One, an instructional session, and a follow-up call at four weeks. The comparison group received a standard pedometer and printed materials with tips for increasing steps. All participants were asked to perform 150 minutes per week of MVPA and walk 10,000 steps per day. A final questionnaire evaluated the assigned intervention, and physical activity was measured for an additional seven days using the ActiGraph GT3X+ accelerometer.
<b>Statistical analysis:</b> List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Data were collected and analyzed according to the intent-to-treat principle. Baseline characteristics were compared using chi-square and t-tests. ActiGraph data were adjusted for number of valid days (95% had 7 valid days; 5% had 5–6 valid days). Baseline-to-16-week physical activity changes were assessed using repeated-measures ANCOVA, adjusted for age and ActiGraph daily wear time to address potential residual confounding.

Timing of measurements: when outcomes were measured; usually baseline and one or more later timesDependent variables: outcomes that are measured or registered;	Participants' physical activity was measured for seven days at baseline and at 16 weeks using the ActiGraph GT3X+. Participants were asked to wear the Fitbit One or pedometer every day throughout the intervention period (112 prescribed days). Physical activity changes in Pedometer Group and Web-Based Tracking Group
variable whose change or different states the researcher wants to understand, explain, or predict	
Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Goal setting process, four-week follow-up call, Fitbit website, printed materials with tips for increasing steps
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	Physical activity according to ActiGraph GT3X+, physical activity changes were adjusted for age and ActiGraph daily wear time
<b>Initial n</b> (e.g. 731 (298 males, 433 females)) <i>Record number that entered</i> <i>study – not the number screened.</i>	51 participants, 0 males and 51 females
<b>Final n</b> (attrition) number of subjects that completed study	51 participants, 0 males and 51 females
Age usually mean or range	53 to 67 years
Ethnicity (if given)	46 participants were non-Hispanic White
<b>Other relevant demographics:</b> <i>demographics describe the</i> <i>population (students, athletes, etc)</i>	32 participants earned a college degree or higher, all participants were comfortable using computers and the internet
Anthropometrics: e.g. were groups same or different on important physical measures (BMI, fitness level)	BMI: 25.7 to 32.7 kg/m2
<b>Location:</b> <i>Where did the study</i> <i>take place? City or country</i>	University of California, San Diego
Summary of Results: Abstract results including quantitative data and statistics. Include statistical significance: P-values, confidence intervals (CI), relative risk (RR), odds ratios (OR), likelihood ratio, number needed to treat, power analysis if available.	At baseline, participants were performing $33\pm56$ min/week of MVPA in bouts $\geq 10$ minutes in length and accumulating $5,866\pm2,195$ steps/day. After the 16-week intervention, the Web-Based Tracking Group increased MVPA by $62\pm108$ min/week (p<.001), MVPA in 10-min bouts by $38\pm83$ min/week (p=0.008), and steps by $789\pm1,979$ (p=0.01), compared to non-significant increases in the Pedometer Group (between-group p-values: 0.11, 0.28 and 0.30, respectively). The Web-Based Tracking Group wore the tracker on 95% of intervention days; 96% reported liking the website, and 100% liked the Fitbit One. <i>Author's Conclusions</i>
Author conclusion: paraphrase that stated by study author in body of the report or abstract	The Fitbit One was well-accepted in this sample of women and was associated with increased physical activity at 16 weeks. By leveraging direct-to-consumer technologies that align with behavior change theories, researchers can strengthen physical activity interventions.

Reviewer comments: Note	Strengths: zero percent attrition, use of baseline and final questionnaires for
strengths and limitations of study;	detailed participant feedback, use of the ActiGraph GT3X+, and use of Fitbit data
identify concerns that affect study	to corroborate adherence
validity and generalizability—	Limitations: small sample size, short duration, and lack of generalizability since all
your comments should be	participants were postmenopausal, overweight/obese women and there were several
italicized)	confounders such as the goal setting process, four-week follow-up call, and Fitbit
	website
	Funding source: this study was funded by NIH (1R03CA168450) and recruitment
	supported by the Athena Breast Health Network. This research was supported by
	the National Cancer Institute (1R03CA168450). The authors have no conflicts of
	interest to report.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet) Cadmus-Bertram, L., Marcus, B., Patterson, R., Parker, B., & Morey, B. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. American Journal of		Y E S	N O	U N C L E	N A
Preventive Medicine, 49(3), 414–418. doi:10.1016/j.amepre.2015.01.020				A	
1. Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some epidemiological studies)	1	Х			
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/ population group would care about?	2	Х			
3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	Х			
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "yes", the report is eligible for design the Evidence Quality Worksheet, depending on answers to the following validity questions.	ation w	ith a	plu	s (+	) on
VALIDITY QUESTIONS		v		N	U N
<b>1. Was the <u>research question</u> clearly stated?</b> <i>This is usually stated at end of the introduction and just before methods section.</i>		E S		0	N A C L E
		Х	ζ.	1	A R
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified? <i>This is often called the treatment and explained in the methods section.</i>	1.1	Х	C .		
<ul> <li>1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?</li> <li>These are sometimes called the endpoints; the results section reports the outcomes, but this information should be in the methods section, too</li> </ul>	1.2	Х	C .		
1.3 Were the target population and setting specified? <i>The target population is group for whom findings may be applicable; look for this in the</i> <i>introduction and in the methods section</i>	1.3	Х	C I		
2. Was the <u>selection</u> of study subjects/patients free from bias?		Y E S		N   O   	UNA CL EA
		Χ	ζ.	]	R
<ul> <li>2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?</li> <li>The authors should give several points about the inclusion/exclusion criteria such as the age range of the subjects, disease condition (like hyperlipidemia) required for inclusion. Exclusion criteria should be listed, too, although some are understood. For example if the ages for inclusion are 18 to 70, the authors will probably not specifically note that children and people over age 70 were excluded. Most of the time, however, subjects may be excluded for certain characteristics such as being pregnant or having some disease (like CHD).</li> <li>2.2 Were criteria applied equally to all study groups?</li> </ul>	2.1	>			
2.2 were britten applied equally to all study gloups?	2.2		7		_
2.5 were health, demographics, and other characteristics of subjects described? There is usually a Table 1 summarizing demographics and characteristics at baseline. Groups are <u>not</u> different if the P-Value is > 0.05. If there has been a previous paper describing the study population, that paper may be referenced and you would need to go back to the original publication to see that Table 1.	2.3				
2.4 Were the subjects/patients in a representative sample of the relevant population? The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion criteria.	2.4			-	X

If there is a large attrition from one group and not others, you would want to look for a reason why, the answer to this question would then be no.       4.5         4.5 If diagnostic test, was decision to perform reference test not dependent on results of test under study?       4.5         The test under study?       4.5         BMI was > 35 but bioimpedance analyzer indicated body far < 30%.       5         5. Was blinding used to prevent introduction of bias?       5.1         5. Was blinding used to prevent introduction of bias?       5.1         5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded       5.1         7.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded       5.1         7.1 The key term is as appropriate?       7         8       8       8         9       9.1 MP on 1 pid levels in hypercholesterolemic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MAT and who was not. Therefore, you would not answer question of blindel for othe detitians and patients to know they were receiving MAT.         9.2. Were data collectors blinded for outcomes assessment? (If outcome is measured to uncome is measured to uncome is measured to the met.)       X         9.3. In cohort study or cross-sectional study, were measureennot outcome is measured on deta study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3         9.3. In cohort study or cross-s				-			
4.5 If diagnostic test, was decision to perform reference test not dependent on results of test under study should be compared to reference test all the time. An example of this might be using a DEXA machine to measure percent body fat only if a subject's BM was >35 but bioimpeduce analyzer indicated body fat < 30%.		If there is a large attrition from one group and not others, you would want to look for a reason why: the answer to this auestion would then be no.					
test under study?       The test under study should be compared to reference test all the time. An example of this might be using a DEXA machine to measure percent body fat only if a subject's BM was >35 but hisingendee analyzer indicated body fat < 30%.		4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5				Х
The test under study should be compared to regreene test of an only of a subject's BMI was > 35 but bioimpedance analyzer indicated body fat < 30%.		test under study?					
Image is a big to Deck indicate analyzer indicated body fat < 30%.		The test under study should be compared to reference test all the time. An example of this					
5. Was blinding used to prevent introduction of bias? <ul> <li>S</li> <li>S</li></ul>		might be using a DEXA machine to measure percent body fat only if a subject s $PMI_{even} > 25$ but biging adapted and an adapted indicated body fat < 200/					
5. Was <u>Difficung</u> used to prevent introduction of Dats?	5	BMI was > 55 but bioimpedance analyzer indicated body jat < 50%.		v	N	U	N
5.1       In intervention study, were subjects, clinicians/practitioners and investigators blinded for treatment group, as appropriate?       X         The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MNT and who was not. Therefore, you would not answer question 5.1 NO. It was appropriate for the dietitians and patients to know they were receiving MNT.         5.2       Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3         5.3       In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       S.3         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       S.4         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       S.5       X         6.       Were intervention study were interventing factors described?       S.5       X       X         6.1       In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       X	5.	was <u>binding</u> used to prevent introduction of blas?		E S	0	N C L E A R	Α
5.1 In intervention study, were subjects, clinicians/practitioners and investigators binded       5.1       X         streatment group, as appropriate?       The key term is as appropriate. For example, in the Lin et al 2008 study, the investigators studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MNT and who was not. Therefore, you would not answer question 5.1 NO. It was appropriate for the distituans and patients to know they were receiving MNT.         5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       5.2         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3       X         5.4 In case control study, were assure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.5       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       A         6.3 Was the intensity and duration of the intervention or exposure factors sufficient to produce a meaningful effect?       6.3       X       A         6.3 Was the intensity and duration of the intervention or exposure fact		5.1. In intermention study, more subjects plinicians/amont/discours and investigations blinded	5 1			X	
The key term is as appropriate. For example, in the Lin et al 2008 study, the investigators studied the effect of M/T on lipid levels in hypercholesterolenic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MNT and who was not. Therefore, you would not answer question 5.1 NO. It was appropriate for the distitutions and patients to know they were receiving MNT.         5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       S.3       X         5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       S.4       X         6.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       S.4       X         6.1 In RCT or other intervention trial, were protocols described?       K       N       K         6.2 In observational study, were intervention study and trait indicans/provider detain?       G.3       X         6.3 Was the intensity and duration of the intervention or exposure factors of procedure and any case and who is a control at the beginning of the study.       S.5       X         6.4. In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       A         6.3 Was the intensity and duratio		5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded to treatment group, <u>as appropriate</u> ?	5.1			Х	
studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was         an RCT, but obviously, the subjects and practitioners knew who was getting MNT         and who was not. Therefore, you would not answer question 5.1 NO. It was         appropriate for the dictitians and patients to know they were receiving MNT.         5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured         using an objective test, such as a lab value, this criterion is assumed to be met.)         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet         study is to have separate people analyzing the data (not the same ones who were         collecting the data).       5.3       X         5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk         factors blinded?         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet         study is to have separate people analyzing the data (not the same ones who were         collecting the data).       X         5.4 In case control study, was case definition explicit and case ascertainment not         influenced by exposure status?        S.4       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any       comparison(s) described in detail? Were intervening factors described?        S.5       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to         produce a meaningful effect?         Use clinical judgment (e.g., 12 weeks is long enough for a dietary intervention to make a         difference in lab values for cholesterol; however, 12 days would not be long         enough)        S.4       X		The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
and who was not. Therefore, you would not answer question 5.1 NO. It was appropriate for the dictitians and patients to know they were receiving MNT.         5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       S.3         5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       X         5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       S.4       X         Establish who is a case and who is a control at the beginning of the study.       S.5       X       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       S.4       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       S.4       X       X         6.4 Was the amount of exposure and, if relevant, subject/patient com		studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					
and who was not. Therefore, you would not answer question 5.1 NO. 11 was       appropriate for the dictitians and patients to know they were receiving MNT.         5.2       Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3       In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       5.3       In cohort study, or to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         6.4       Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervention fraid. Were intervention factors study, were interventions, study settings, and clinicians/provider described?       6.1       X         6.1       In RCT or other intervention rial, were protocols described for all regimens studied?       6.1       X         6.2       In was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         0.4       Was the intensity and duration of the intervention or exposure factor sufficient to make a difference in lab values for cholesterol; however, 12 days would		an RCT, but obviously, the subjects and practitioners knew who was getting MNT					
appropriate for the dictitians and patients to know they were receiving MAL.       Image: Comparison of the dictitians and patients to know they were receiving MAL.         5:2       Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)       S.2         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       S.3       In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       S.4       X         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       S.4       X         Establish who is a case and who is a control at the beginning of the study.       S.5       X       X         6.       Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       S.7       X         6.1       In RCT or other intervention trial, were protocols described for all regimens studie?       6.1       X       K         6.2       In observational study, were tay therevention or exposure factor sufficient to produce a meaningful effect?       S.8       C		and who was not. Therefore, you would not answer question 5.1 NO. It was					
5.2       Were data collectors binded for outcomes assessment? (If outcome is measured to be met.)       5.2       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3       In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       5.3       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         5.5       In diagnostic study, were test results blinded to patient history and other test results?       5.5       X         6.       Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       5       S       X         6.1       In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       E         6.3       Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       K       K       K         6.4       Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X <td< td=""><td></td><td>appropriate for the dictitians and patients to know they were receiving MN1.</td><td>5.2</td><td></td><td></td><td>v</td><td></td></td<>		appropriate for the dictitians and patients to know they were receiving MN1.	5.2			v	
Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3       S.3       X         5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       5.3       X       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       5.5       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X         6.2 In observational study, were intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         16.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         7       6.6 Were exta or unplanned treatments other therapies) described?       6.5       X         6.5 Were control study, were intervention the same		5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an abjective test, such as a lab value, this criterion is assumed to be met.)	5.2			А	
Answer yes, 0 that test was taste to measure an outcome. A method of othing a thet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.3       X         5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       5.3       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         5.5 In diagnostic study, were test results blinded to patient history and other test results?       5.5       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       6.1       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studie?       6.1       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       X         0.5.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X       X         6.5 Were exit or unplanned treatments		Answer yes, if a lab test was used to measure an outcome. A method of blinding a dist					
Solution is a construction of the intervention of the i		Answer yes, if a lab lesi was used to measure an outcome. A method of blinding a diel					
5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?       5.3       X         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).       5.4       X         5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         Establish who is a case and who is a control at the beginning of the study.       5.5       X       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N       N       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       X         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.3       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       S.4       X       X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X       X       K </td <td></td> <td>collecting the data)</td> <td></td> <td></td> <td></td> <td></td> <td></td>		collecting the data)					
In actors binded?         Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data).         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?         Establish who is a case and who is a control at the beginning of the study.       5.5         5.5       In agnostic study, were test results blinded to patient history and other test results?       5.5         6.       Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       5.1         6.1       In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X         6.2       In observational study, were intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X         6.4       Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         6.4       Was the amount of e.g., ancillary treatments other therapies) described?       6.5       X         6.6       Were extra or unplanned treatments described?       6.6       X		5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk $f_{1}$	5.3				Х
Answer yes, if a tab lest was used to medsure an olifcome. A method of bilning a dief       study is to have separate people analyzing the data (not the same ones who were collecting the data).         5.4       In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         Establish who is a case and who is a control at the beginning of the study.       5.5       X       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       X $\frac{k}{R}$ 6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       Image: Control at the begin of the intervention or exposure factor sufficient to produce a meaningful effect?       X $\frac{k}{R}$ 6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       Image: Control and the box of the study would not be long enough)         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X       Image: Control and the treatment last? Did the patient follow directions?       6.5       X       Image: Control and the study would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall       6.6       X		factors blinded?					
Staty is to have separate people analyzing the data (not the same ones who were collecting the data). <ul> <li>S.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>S.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>S.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>S.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>S.5 In diagnostic study, were test results were factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>S.5 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>G.1 In RCT or other intervention of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> <li>A was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>A was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>A was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>A was the anount of exposure and the same time as the diet therapy)</li> <li>A was the anount of exposure and the same time as the diet therapy)</li> <li>A was the anount of exposure and treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a moble</li></ul>		Answer yes, if a lab lest was used to measure an outcome. A method of blinding a diel					
5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?       5.4       X         5.5 In diagnostic study, were test results blinded to patient history and other test results?       5.5       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N       X         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       X         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.2       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X       X         6.6 Were extra or unplanned treatments described?       6.6       X       X         6.6 Were extra or unplanned treatment described?       6.6       X       X		collecting the data)					
5.4 In clase control study, was clase case and who is a control at the beginning of the study.       5.4 X         5.5 In diagnostic study, were test results blinded to patient history and other test results?       5.5 X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1 X       X       X         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.2 X       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3 X       Image: Single study sets is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4 X       Image: Single study sets is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.5 X       Image: Single study sets is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       Image: Single study sets is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       Image: Single study sets is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       Image: Single study sets is long enough for a diet		5.4 In case control study, was case definition explicit and case ascertainment not	54				X
Establish who is a case and who is a control at the beginning of the study.       5.5       X         5.5 In diagnostic study, were test results blinded to patient history and other test results?       5.5       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N       V       N       A         6. 1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       X       A         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.2       X       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       Image: Studie of the study o		influenced by exposure status?	J. <b>T</b>				Λ
5.5 In diagnostic study, were test results blinded to patient history and other test results?       5.5       X         6. Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N       N </td <td></td> <td>Establish who is a case and who is a control at the beginning of the study.</td> <td></td> <td></td> <td></td> <td></td> <td></td>		Establish who is a case and who is a control at the beginning of the study.					
6. Were intervention/herapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?       Y       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V       N       V <t< td=""><td></td><td>5.5 In diagnostic study, were test results blinded to patient history and other test results?</td><td>5.5</td><td>1</td><td></td><td></td><td>Х</td></t<>		5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5	1			Х
comparison(s) described in detail? Were intervening factors described?       Image: Solution of the solution of the intervening factors described?       Image: Solution of the intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       Image: Solution of the intervention of the intervention of the intervention of the patient follow directions?       Image: Solution of the intervention of the intervention of the intervention of the intervention of the interventions?         6.4       Was the amount of exposure and, if relevant, subject/patient compliance measured?       Image: Solution of the intervention of the intervention of the intervention of the interventions?       Image: Solution of the intervention of the intervention of the intervention of the intervention of the interventions?       Image: Solution of the intervention of the intervention of the intervention of the intervention of the intervention?       Image: Solution of the intervention of the intervention of the intervention?       Image: Solution of the intervention of the intervention of the intervention?       Image: Solution of the intervention of the intervention of the intervention?       Image: Solution of the intervention of the interventinterventintervention of the intervention of the interv	6.	Were intervention/therapeutic regimens/exposure factor or procedure and any		Y	N	U	N
6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X       Image: Construction of the intervention of the intervention of the intervention of exposure factor sufficient to produce a meaningful effect?       6.3       X       Image: Construction of the intervention of the intervention of the intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       Image: Construction of the intervention of the intervention of the intervention of the intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       Image: Construction of the intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       Image: Construction of the intervent of the intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X       Image: Construction of the intervent of the intervention of the interventions?       6.5       X       Image: Construction of the intervent of the inthe intervent of the intervent of the intervent of the		comparison(s) described in detail? Were intervening factors described?		S	0	C N	А
A       X       A         6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.2       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       Intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough       6.4       X       Interventions?         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X       Interventions?         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X       Intervention?         6.6 Were extra or unplanned treatments described?       6.6       X       X         7       7       6.6       X       X         8       9       9       10       10       10         9       10       10       10       10       10       10         10       10       10       10       10       10       10       10         10       10       10       10       10       10       10       10       10       10       10       1						L E	
6.1 In RCT or other intervention trial, were protocols described for all regimens studied?       6.1       X         6.2 In observational study, were interventions, study settings, and clinicians/provider described?       6.2       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X       Image: Clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X         6.6 Were extra or unplanned treatments described?       6.6       X         7. He text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall       6.4       X				Х		A R	
6.2 In observational study, were interventions, study settings, and clinicians/provider       6.2       X         6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X         6.6 Were extra or unplanned treatments described?       6.6       X         7 the text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall       6.6       X		6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1	Х			
described?       6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3 X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4 X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4 X         How long did the treatment last? Did the patient follow directions?       6.5 X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5 X         (e.g. were patients on lipid-lowering meds at the same time as the diet therapy)       6.6 X         6.6 Were extra or unplanned treatments described?       6.6 X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall		6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2				Х
6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?       6.3       X         Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4       X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         How long did the treatment last? Did the patient follow directions?       6.5       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X         6.6 Were extra or unplanned treatments described?       6.6       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall       6.6       X		described?					
produce a meaningful effect?       Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)       6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4 X         6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4 X          6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5 X          6.6 Were extra or unplanned treatments described?       6.6       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       6.6       X		6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough) <ul> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?</li> <li>6.5 Were extra or lipid-lowering meds at the same time as the diet therapy)</li> <li>6.6 Were extra or unplanned treatments described?</li> <li>6.6 Were extra or unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"</li> <li>answer is not a problem overall</li> </ul>		produce a meaningful effect?					
difference in lab values for cholesterol; however, 12 days would not be long enough) <ul> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> <li>6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?</li> <li>6.5 Were extra or unplanned treatments described?</li> <li>6.6 Were extra or unplanned treatments described?</li> <li>6.6 Were extra or unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no" answer is not a problem overall</li> </ul>		Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
enough)       6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4 X         How long did the treatment last? Did the patient follow directions?       6.4 X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5 X         (e.g. were patients on lipid-lowering meds at the same time as the diet therapy)       6.6 X         6.6 Were extra or unplanned treatments described?       6.6 X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       X		difference in lab values for cholesterol; however, 12 days would not be long					
6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?       6.4       X         How long did the treatment last? Did the patient follow directions?       6.4       X         6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X         (e.g. were patients on lipid-lowering meds at the same time as the diet therapy)       6.5       X         6.6 Were extra or unplanned treatments described?       6.6       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       X		enough)					
How long did the treatment last? Did the patient follow directions?       6.5         6.5       Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5       X         (e.g. were patients on lipid-lowering meds at the same time as the diet therapy)       6.6       X         6.6       Were extra or unplanned treatments described?       6.6       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       6.6       X		6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.4	Х			
6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?       6.5 X         (e.g. were patients on lipid-lowering meds at the same time as the diet therapy)       6.6 X         6.6 Were extra or unplanned treatments described?       6.6 X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       6.6 X		How long did the treatment last? Did the patient follow directions?	6.5				
6.6       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       6.6		6.5 were co-interventions (e.g., ancillary treatments other therapies) described?	6.5	X			
0.0 Were extra of unplanned treatments described?       0.0       X         The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"       0.0       X		(e.g. were patients on lipid-lowering meas at the same time as the alet therapy)	6.6				v
discussion section. It is likely there were no unplanned treatments, so a "no"	1	o.o were extra or unplanned treatments described?	0.0				Å
answer is not a problem overall	1	i ne iexi muy noi describe any unplannea irealments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"					
CONTRACT AN INTELLE DECEMBER OF VERY ALL	1	assuer is not a problem overall					
67 Was the information for 64 65 66 and 67 assessed the same way for all groups?	<u> </u>	(7 Westhe information for (A, (5, (), and (7) assessed the same may for all around?)	67	x	╞╴┤		
The set of the information for 0.1, 0.5, 0.0 and 0.7 abbedded the build way for an groups: 0.7 A		o / was the information for 0.4 o 5 o 6 and 6 / assessed the same way for all orouns /	· · · ·				

	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8				Х
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E A	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X		R	
	7.2 Were nutrition measures appropriate to question and outcomes of concern? Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul><li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li><li>Check that surveys were validated.</li></ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability.</i>	7.5	Х			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7	Х			
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	X		ĸ	 
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)	8.3	Х			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4	X			
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5	X			
	8.6 Was clinical significance as well as statistical significance reported? <i>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was</i> <i>reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical</i>	8.6	X			

		-		1	
significance (P-value) and clinical significance (compare to standard of $< 200$					
mg/do for normal cholesterol). A problem can occur when only statistical					
significance is reported. Reducing cholesterol from 300 to 250 might be statistically					
significant, but clinically it is still abnormal.	0.7				37
8.7 If negative findings, was a power calculation reported to address type 2 error?	8.7				Х
Type II ( $\beta$ error is a false negative that happens when the investigators fail to reject the					
<u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say					
something like "a sample size of $n=xx$ is needed to provide 80% power."		V	N	T	N
9. Are <u>conclusions</u> supported by results with biases and limitations taken into		ч Е	0	N	A
consideration?		s		C L	
				E	
		Х		A R	
9.1 Is there a discussion of findings?	9.1	Х			
Answer yes or no.					
9.2 Are biases and study limitations identified and discussed?	9.2	Х			
This will be in the discussion of finding section that follows the results					
10. Is bias due to study's funding or sponsorship unlikely?		Y	N O	U N	N A
Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		ŝ	Ŭ	C	
				L E	
		Х		A R	
10.1 Were sources of funding and investigators' affiliations described?	10.1	X			
• Look just under the abstract. or					
• The funding may be acknowledged at the end of the paper					
<ul> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>					
10.2 Was there no apparent conflict of interest?	10.2	X			
If an investigator is testing a niece of equipment process or drug that s/he developed it	10				
<i>could be a conflict of interest</i>					
SVMBOL					
MINUS/NEGATIVE (-)					_
If most (six or more) of the answers to the above validity questions are "no" the report should	he desior	nated	with	n a	
minus (-) symbol on the Evidence Quality Worksheet	ee acoisis.	uicu			
NEUTRAL (a)					
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is except	tionally s	trono	the	2	
report should be designated with a neutral ( $\alpha$ ) symbol on the Evidence Quality Worksheet			,		
PLUS/POSITIVE (+)					

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Citation	Cadmus-Bertram, L., Marcus, B., Patterson, R., Parker, B., & Morey, B. (2015). Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. American Journal of Preventive Medicine, 49(3), 414–418. doi:10.1016/j.amepre.2015.01.020
Study Design	Randomized controlled trial
Class	Α
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\bigcirc$ (Neutral)
Research Purpose	To evaluate, within a randomized controlled trial, the feasibility and preliminary efficacy of integrating a Fitbit tracker and website into a physical activity intervention for postmenopausal, overweight/obese women
Inclusion Criteria	Postmenopausal women, overweight/obese, perform less than 60 minutes per week of moderate-to-vigorous intensity physical activity (MVPA), could exercise safely, were regular internet users, and owned a tablet/computer
Exclusion Criteria	Women who responded, "somewhat uncomfortable," "very uncomfortable," "somewhat do not enjoy," or "very much do not enjoy" on baseline questionnaire
Description of Study Protocol	Recruitment: Unclear, supported by the Athena Breast Health Network Design: Participants were randomized to either a Fitbit or pedometer- based intervention group to determine whether the Fitbit One increased physical activity more than the pedometer in postmenopausal, overweight/obese women Blinding used (if applicable): Unclear Intervention (if applicable): Participants received a baseline questionnaire, anthropometric measurements, and physical activity was measured for seven days using the ActiGraph GT3X+ accelerometer. Then, participants were randomized to a 16-week web-based self- monitoring intervention (N=25) or a comparison group (N=26). Participants in the Web-Based Tracking Group received a Fitbit One, an instructional session, and a follow-up call at four weeks. The comparison group received a standard pedometer and printed materials with tips for

	week of MVPA and walk 10,000 steps per day. A final questionnaire
	evaluated the assigned intervention, and physical activity was measured
	for an additional seven days using the ActiGraph GT3X+ accelerometer.
	Statistical Analysis: Data were collected and analyzed according to the
	intent-to-treat principle. Baseline characteristics were compared using
	chi-square and t-tests. ActiGraph data were adjusted for number of valid
	days (95% had 7 valid days; 5% had 5-6 valid days). Baseline-to-16-
	week physical activity changes were assessed using repeated-measures
	ANCOVA, adjusted for age and ActiGraph daily wear time to address
	potential residual confounding.
	Timing of Measurements: Participants' physical activity was measured for
	seven days at baseline and at 16 weeks using the ActiGraph GT3X+.
	Participants were asked to wear the Fitbit One or pedometer every day
	throughout the intervention period (112 prescribed days).
Data Collection	Dependent Variables: Physical activity changes in Pedometer Group and
Summary	Web-Based Tracking Group
	Independent Variables: Goal setting process, four-week follow-up call,
	Fitbit website, printed materials with tips for increasing steps
	Control Variables: Physical activity according to ActiGraph GT3X+,
	physical activity changes were adjusted for age and ActiGraph daily wear
	time
	Initial: 51 (0 Males 51 Females)
	Attrition (final N): 51
	Age: 53 to 67 years
Description of	Age: 53 to 67 years Ethnicity: 46 participants were non-Hispanic White
Description of Actual Data Sample	Age: 53 to 67 years Ethnicity: 46 participants were non-Hispanic White Other relevant demographics: 32 participants earned a college degree or
Description of Actual Data Sample	Age: 53 to 67 years Ethnicity: 46 participants were non-Hispanic White Other relevant demographics: 32 participants earned a college degree or higher, all participants were comfortable using computers and the internet
Description of Actual Data Sample	Age: 53 to 67 years Ethnicity: 46 participants were non-Hispanic White Other relevant demographics: 32 participants earned a college degree or higher, all participants were comfortable using computers and the internet Anthropometrics: BMI: 25.7 to 32.7 kg/m2

	Key Findings: At baseline, participants were performing 33±56 min/week
	of MVPA in bouts $\geq 10$ minutes in length and accumulating 5,866±2,195
	steps/day. After the 16-week intervention, the Web-Based Tracking
	Group increased MVPA by 62±108 min/week (p<.001), MVPA in 10-min
	bouts by 38±83 min/week (p=0.008), and steps by 789±1,979 (p=0.01),
Summary of	compared to non-significant increases in the Pedometer Group (between-
Kesuits	group p-values: 0.11, 0.28 and 0.30, respectively). The Web-Based
	Tracking Group wore the tracker on 95% of intervention days; 96%
	reported liking the website, and 100% liked the Fitbit One.
	Other Findings:
	The Fitbit One was well-accepted in this sample of women and was
Author	associated with increased physical activity at 16 weeks. By leveraging
Conclusion	direct-to-consumer technologies that align with behavior change theories,
	researchers can strengthen physical activity interventions.
	Strengths: zero percent attrition, use of baseline and final questionnaires
	for detailed participant feedback, use of the ActiGraph $GT3X+$ , and use of
	Fitbit data to corroborate adherence
Reviewer	Limitations: small sample size, short duration, and lack of generalizability
Comments	since all participants were postmenopausal, overweight/obese women and
	there were several confounders such as the goal setting process, four-
	week follow-up call, and Fitbit website
	This study was funded by NIH (1R03CA168450) and recruitment
	supported by the Athena Breast Health Network. This research was
Funding Source	supported by the National Cancer Institute (1R03CA168450). The
	authors have no conflicts of interest to report.

## Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
4	<b>Positive</b> – Indicates that the report has clearly addressed issues of
÷	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.

exceptionally week
--------------------

		Select a rating from the drop-down menu ↓			
Re	elevance Questions				
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Yes		
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes		
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	Yes		
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes		

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Va	alidity Questions		
1.	Was the <u>research question</u> clearly stated?	1	Yes
	1.1. Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3. Were the target population and setting specified?	1.3	Yes
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Yes
	2.1. Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes
	without omitting criteria critical to the study?	2.2	Yes
	<ol> <li>Were criteria applied equally to all study groups?</li> <li>Were health, demographics, and other characteristics of subjects described?</li> </ol>		Y N
	2.4. Were the subjects/patients a representative sample of the relevant	2.3	Yes
	population?	2.4	Unclear
3.	Were study groups comparable?	2	Vac
	3.1. Was the method of assigning subjects/patients to groups described and	5	105
	unbiased? (Method of randomization identified if RCT)	3 1	Vac
	demographics) similar across study groups at baseline?	5.1	105
	<ol> <li>Were concurrent controls used? (Concurrent preferred over historical controls.)</li> </ol>	3.2	Yes
	3.4. If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by usin	g 3.3	Yes
	<ul><li>appropriate adjustments in statistical analysis?</li><li>3.5. If case control study, were potential confounding factors comparable for cases</li></ul>	3.4	N/A
	and controls? (If case series or trial with subjects serving as own control, this criterion is not applicable. Criterion may not be applicable in some cross-sectional studies.)	3.5	N/A
	3.6. If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	N/A
4.	Was method of handling withdrawals described?	4	N/A

	4.1. Were follow up methods described and the same for all groups?	4.1	N/A
	4.2. Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow	12	N/A
	up, attrition rate) and/or response rate (cross-sectional studies) described for each group? (Follow up goal for a strong study is 80% )	4.2	IN/A
	4.3. Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Yes
	4.4. Were reasons for withdrawals similar across groups	4.4	N/A
	4.5. If diagnostic test, was decision to perform reference test not dependent on	45	N/A
-	results of test under study?	1.5	1 1/1 1
5.	5.1 In intervention study, were subjects, clinicians/practitioners, and investigators	5	Unclear
	blinded to treatment group, as appropriate?		
	5.2. Were data collectors blinded for outcomes assessment? (If outcome is	5.1	Unclear
	measured using an objective test, such as a lab value, this criterion is assumed	5.2	Unclear
	to be met.)	5.2	Olicical
	5.3. In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	5.3	N/A
	5.4. In case control study, was case definition explicit and case ascertainment not	54	N/A
	influenced by exposure status?	5.1	1 1/2 1
	results?	5.5	N/A
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any	6	Yes
	comparison(s) described in detail? Were <u>intervening factors</u> described?	6.1	X7
	6.1. In RCI or other intervention trial, were protocols described for all regimens studied?	6.1	Yes
	6.2. In observational study, were interventions, study settings, and	6.2	N/A
	clinicians/provider described?	63	Ves
	6.3. Was the intensity and duration of the intervention or exposure factor sufficient	0.5	105
	to produce a meaningful effect?	6.4	Yes
	measured?	6.5	Yes
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.6	N/A
	6.7 Was the information for 6.4.6.5 and 6.6 assessed the same way for all groups?	67	V
	6.8. In diagnostic study, were details of test administration and replication	0./	res
	sufficient?	6.8	N/A
7.	Were <u>outcomes</u> clearly defined and the <u>measurements valid and reliable</u> ?	7	Yes
	7.1. Were primary and secondary endpoints described and relevant to the	7.1	Yes
	7.2. Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A
	7.3. Was the period of follow-up long enough for important outcome(s) to occur?	73	N/A
	7.4. Were the observations and measurements based on standard, valid, and	7.5	Vac
	reliable data collection instruments/tests/procedures?	7.4	res
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.7. Were the measurements conducted consistently across groups?	7.6	Yes
	······································	7.7	Yes
8.	Was the statistical analysis appropriate for the study design and type of outcome indicators?	8	Yes
	8.1. Were statistical analyses adequately described the results reported	8.1	Yes
	appropriately? 8.2. Were correct statistical tests used and assumptions of test not violated?	8.2	Yes

8.3.	Were statistics reported with levels of significance and/or confidence intervals?	8.3	Yes
8.4.	Was "intent to treat" analysis of outcomes done (and as appropriate, was there		
	an analysis of outcomes for those maximally exposed or a dose-response	8.4	Yes
		85	Ves
8.5.	Were adequate adjustments made for effects of confounding factors that	0.5	105
	might have affected the outcomes (e.g., multivariate analyses)?	06	Var
8.6.	Was clinical significance as well as statistical significance reported?	0.0	105
07	If nogative findings, was a newer calculation reported to address type 2 error?	07	NT/A
0.7.	in negative infumes, was a power calculation reported to address type 2 error:	8.7	IN/A
9. Are <u>co</u>	nclusions supported by results with biases and limitations taken into	9	Yes
consid	leration?	0.1	Var
91	Is there a discussion of findings?	9.1	res
0.2	Are biases and study limitations identified and discussed?	92	Yes
9.2.	Are blases and study limitations identified and discussed?	··-	105
10. Is bias	due to study's <u>funding or sponsorship</u> unlikely?	10	Yes
10.1	. Were sources of funding and investigators' affiliations described?	10.1	Ves
10.2	Was there no apparent conflict of interest?	10.1	105
10.2		10.2	Yes
NATINITIC AL			

#### MINUS/NEGATIVE (-)

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

#### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

## **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Chum, J., Kim, M., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., Samaan, Z. (2017). Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. <i>Evidence-Based Mental Health</i> , <i>20</i> (4), 128–133. doi:10.1136/eb-2017-102763
<b>Study design:</b> Use algorithm – <i>RCT, cohort, etc</i>	Randomized controlled trial
Study Class (A,B,C,D)	Α
<b>Research Quality Rating</b> This rating tells if the research design is good (+), bad (-) or neutral (Ø) This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).	NEUTRAL (0)
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the research question being investigated in the study?)	To understand patients' perceived benefit from the Fitbit One and explore themes associated with patient experiences. To compare perceived benefit, patient factors, Fitbit usage, and Beck's Depression Inventory (BDI) scores.
Inclusion criteria: requirements for study eligibility	Patients 18 years or older with major depressive disorder were approached to participate in the Behavioural Activation Group Program in Patients with Depression (BRAVE) study. All patients with depressive disorder receiving treatment for depression, including pharmacotherapy and psychotherapy were eligible to participate. Among the 87 participants who completed the BRAVE study, 36 participants completed interviews.
<b>Exclusion criteria</b> (conditions that make individual ineligible)	Inability to provide written informed consent, inability to understand written and spoken English, and having a primary diagnosis other than depression
Recruitment	Patients who were attending or referred for an assessment at the mood disorders outpatient clinic
Blinding used: some of the persons involved are prevented from knowing certain information that might lead to conscious or unconscious bias on their part, invalidating the results	Research assistants collected, recharged, and synced data from each participants' Fitbit One to their online account. Personalized printouts were shared with participants to minimize data manipulation. Research assistants also conducted individual, anonymized interviews with patients who completed the BRAVE study.
<b>Description of study protocol</b> <i>What happened in the study?</i>	During the BRAVE study, patients were asked to carry a Fitbit One for 18-weeks. Interviews were conducted with 36 patients who completed the BRAVE study. Thematic analyses were conducted on the interviews and exploratory quantitative analyses were conducted on patient characteristics, Fitbit usage, steps recorded, perceived benefit, and BDI scores.
Intervention: Describe interventions, regimens, risk factors, or procedures studied.	Fitbit One activity trackers were provided to all patients allocated to the behavioral activation (BA) group at the beginning of the BRAVE study, and email accounts were created for each participant. The patients were encouraged to carry the Fitbit One at all times throughout the 18 weeks to track their physical activity. At the beginning of each BA session, Fitbits were collected, recharged, and synced to the online accounts of their respective holders. The participants' weekly activity data were then exported from the Fitbit website, and a personalized printout was generated including summaries of step count, stairs climbed, and distance travelled. The printouts were shared with participants as they did not have access to the online accounts to minimize data manipulation. However, the Fitbit One can show the number of steps taken daily, which can be seen by the patient. Qualitative data

	were collected through individual, semi-structured interviews, which aimed to explore patients' experiences with their BA treatment and their Fithit use
	Additional data such as baseline BDI scores, pretherapy goals, familiarity with
	technology, and other patient characteristics were collected through questionnaires
	facilitated by a clinician during the study intake. BDI scores were collected weekly
	throughout the study including at completion. Exploratory analyses were
	conducted to examine how perceived benefit related to participants' baseline
	characteristics, including age, BDI scores, pretherapy physical activity goals, and
	completion and Fitbit usage.
Statistical analysis: List tests,	The control and intervention groups were examined through chi-square tests. To
significance level set a priori	examine the relationships between baseline characteristics and Fitbit use,
( $\alpha$ =0.05; include intent to treat	independent t-tests were used. Between group differences were also calculated
analysis if applicable; note if	through independent t-tests. Pearson's correlation was used to compare Fitbit
there is Power analysis.	usage with age and baseline BDI scores.
Timing of measurements: when	I rained clinicians delivered BA to participants randomized to receive the intervention up to twice nor weak over 18 weaks for a total of 28 coscions.
baseline and one or more later	Intervention up to twice per week over 18 weeks, for a total of 28 sessions.
times	neriod Individual interviews lasted 30 to 60 minutes and were finished within two
umes	months of participants' completion of the BA program
Dependent variables: <i>outcomes</i>	Patients' perceived benefits. Fitbit usage. BDI scores
that are measured or registered;	1 , 5,
variable whose change or	
different states the researcher	
wants to understand, explain, or	
predict	
Independent variables	BA topics, such as value assessment, goal setting, leisure education
(intervention or procedure; this	
variable whose effect upon the	
dependent variable one is trying	
to understand)	
Control Variables	Receiving treatment for depression as usual, including pharmacotherapy and
Examples: 1) multivariate logistic	psychotherapy
regression controlled for age,	
BMI, albumin;	
2) usual care; 3) isocaloric diet,	
<i>etc.</i>	26 martin and a 10 martin and 10 females
Initial n (e.g. $/31$ (298 males, $/32$ famplas))	36 participants, 18 males and 18 females
Record number that entered	
study – not the number screened	
<b>Final n</b> (attrition)	36 participants, 18 males and 18 females
number of subjects that completed	
study	
Age usually mean or range	Mean age: 53 years
Ethnicity (if given)	Unclear
Other relevant demographics:	Mean BDI score: 36.27 (scale of 0 to 63)
demographics describe the	23 participants had prior smartphone use
population (students, athletes, etc)	20 participants attended a college, university, or higher education
arouns same or different on	IVICAII DIVII. 50.10 Kg/III2
important physical measures	
(BML fitness level)	
12111, Junobb 10101	

Location: Where did the study	Mood disorders outpatient clinic in Hamilton, Ontario, Canada
take place? City or country	
Summary of Results: Abstract results including quantitative data and statistics. Include statistical significance: P-values, confidence intervals (CI), relative risk (RR), odds ratios (OR), likelihood ratio, number needed to treat, power analysis if available.	23 patients found the Fitbit One to be helpful for their physical activity. Themes of positive experiences included 27 patients who reported self-awareness, three patients who reported peer motivation, five patients who reported goal-setting opportunities, and 13 patients who reported enjoying using the Fitbit One. Themes of negative experiences included 12 patients who reported inconvenience, 11 patients who reported inaccuracies, two patients who reported discouragement, and 10 patients who reported disinterest. No significant differences were found in age (p=0.72), baseline BDI scores (p=0.44), percent change in BDI scores (p=0.283), pretherapy physical activity goals (p=0.549), and smartphone use (p=0.825) between those who did and did not find the Fitbit One helpful. However, there was a significant relationship between total Fitbit One usage and perceived benefit. The mean number of weeks of Fitbit One use for those who found the Fitbit helpful was 18 57 (SD 1 21) and 12 27 (SD 5 76) weeks for those who did not (p<0.001)
	Author's Conclusions
Author conclusion: paraphrase	Of the 36 patients who underwent the BRAVE study and completed interviews the
that stated by study author in	of the 50 patients who under whit the DRAVE study and completed interviews, the majority $(64\%)$ found the Eithit One to be helpful identifying their Eithit use as a
had stated by study duthor in	factor in improving their physical activity level. Many positive themes were
body of the report of abstract	appeardent with current literature: however, notients also reported pagetive appear
	concordant with current merature, nowever, patients also reported negative aspects
	that may affect use. Interestingly, there was no significant relationship between
	perceived benefit of the Fitbit and percent change in BDI scores, which contradicts
	previous literature supporting Fitbit use in treating depression. These findings
	suggest that the Fitbit One may be useful for patients with varying characteristics.
	Clinicians and researchers should consider both strengths and limitations of activity
	trackers when implementing them to motivate patients with depression.
Reviewer comments: Note	Strengths: zero percent attrition, use of baseline data and final interviews for
strengths and limitations of study:	detailed participant feedback, and use of Fitbit data to corroborate adherence
identify concerns that affect study	Limitations: small sample size compared to the number of participants who
validity and generalizability	completed the <i>BRAVE</i> study difficulty interpreting exploratory findings given that
your comments should be	the study was undernowered to test affectiveness, and there were several
italicized)	approximate state and a second and the second section and approximate and approximate and approximate and approximate and approximate and approximate approximate and approximate approxim
пипст2еи)	contounders such as comonieu DA miervention, study setting, and participants
	Funding source: supported by Canadian Institutes of Health Research (CIHR)
	(Randomised Controlled Trials: Mentoring, code number 201303MTP-303860-
	182743) to conduct the BRAVE study. CIHR mentorship award (2013–2015) is to
	support research time. The authors declared no competing interests.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)		Y	N	U	N
Chum, J., Kim, M., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., Samaan, Z. (2017).		E S	0	N C	А
Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study.				L	
Evidence-Based Mental Health 20(4) 128–133 doi:10.1136/eb-2017-102763				A	
1 Would implementing the studied intervention or precedure (if found successful) result in	1	v		R	
1. would implementing the studied intervention of procedure (in found successful) result in	1	Λ			
improved outcomes for the patients/clients/population group? (Not Applicable for some					
2 Did the end of the e	2	v			
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	Χ			
population group would care about?	2		17		
3. Is the focus of the intervention or procedure (independent variable) or topic of study a	3		Х		
common issue of concern to dietetics practice?					
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	X			
If the answers to all of the above relevance questions are "yes", the report is eligible for design	ation w	ith a	plu	s (+)	on
the Evidence Quality Worksheet, depending on answers to the following validity questions.					
VALIDITY QUESTIONS					
1. Was the <u>research question</u> clearly stated?		Y	I	N U D N	N A
This is usually stated at end of the introduction and just before methods section.		ŝ		Ċ	
				E	
		Х	2	A	
1.1 Was the enceific intervention(a) or precedure (independent verichle(a)) identified?	1 1	v	,		
This is a few collection intervention(s) of procedure (independent variable(s)) identified?	1.1	Δ	•		
This is often called the treatment and explained in the methods section.	1.0		,		_
1.2 was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	X	•		
These are sometimes called the endpoints; the results section reports the outcomes, but					
this information should be in the methods section, too	1.0		-		
1.3 Were the target population and setting specified?	1.3	X			
The target population is group for whom findings may be applicable; look for this in the					
introduction and in the methods section		v	,	JT	N
2. Was the <u>selection</u> of study subjects/patients free from bias?		E	ĺ		A
		s			
			-	E	
		X		R	
2.2 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1	Х			
diagnostic or prognosis criteria), and with sufficient detail and without omitting					
criteria critical to the study?					
The authors should give several points about the inclusion/exclusion criteria such as the					
age range of the subjects, disease condition (like hyperlipidemia) required for					
inclusion. Exclusion criteria should be listed, too, although some are					
understood. For example if the ages for inclusion are 18 to 70, the authors will					
probably not specifically note that children and people over age 70 were					
excluded. Most of the time, however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD).					
2.3 Were criteria applied equally to all study groups?	2.2	X	2		
2.4 Were health, demographics, and other characteristics of subjects described?	2.3	Х	2		
There is usually a Table 1 summarizing demographics and characteristics at baseline.					
Groups are <u>not</u> different if the P-Value is $> 0.05$ . If there has been a previous					
paper describing the study population, that paper may be referenced and you					
would need to go back to the original publication to see that Table 1.	1				
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4	1		Z	Κ
The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may					
only say that the patients came from the same clinic from people who met the inclusion					
criteria.	1				

3. \	Were <u>study groups comparable</u> ?		Y E	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		s	Ŭ	Ĉ	
	Groups are <u>not</u> different if the P-Value is $> 0.05$ .				E	
					AR	
					X	
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1			X	
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design: or assigned by computer-generated random numbers).					
	Look for instances that show bias: for example I once read a study where patients					
	were randomized to receive liquid energy supplements: however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status prognostic factors and other factors (e.g.	32			X	
	demographics) similar across study groups at baseline? See Table I for this - there	5.2				
	should be no significant differences across study groups in an intervention study					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	33		v		
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a	5.5		Λ		
	most RC15 use a control group: that is an example of a historical control. That is not					
	prior study as a control group, that is an example of a historical control. That is not					
	us strong a research design as use of concurrent control group. A crossover study					
	2.4. If expert study or energy sectional study, were groups comparable on important	2.4	-			v
	5.4 If conort study of closs-sectional study, were groups comparable on important	3.4				Λ
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a conort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.	2.5				
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6				Х
	reference standard (e.g. "gold standard")?					
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4.	Was method of handling <u>withdrawals</u> described?		Y E	N O	U N	N A
			s		C	
					E	
					A	Х
	4.1 Were follow up methods described and the same for all groups?	4.1			ĸ	Х
	4.2 Was the number characteristics of withdrawals (i.e. dronouts lost to follow up	4.2		1		X
	attrition rate) and/or response rate (cross sectional studies) described for each group?	т.2				Λ
	(Follow up goal for a strong study is 80 %)					
	(Follow up goal for a strong study is 80 70.) This should be found in the results section. If there is attrition $> 20\%$ it is important to					
	note that on the worksheet (as a note in the results section or in the results					
	comments at the very bottom)					
	4.2 Ware all approlled subjects/patients (in the original sample) accounted for?	1 2		v		
	4.5 were an enfonce subjects/patients (in the original sample) accounted 101?	4.3	1			
	<i>I his information is often presented in a figure with # recruited, # enrolled (INIS IS Initial</i>		1	1		
	<i>iv)</i> , <i># remaining at end of study period (final IV)</i> . Sometimes the reasons that subjects with draw on work dropped is given in the forward or in the text (course to be set in the set of the set o		1	1		
	4.4 Ware reasons for with drownla similar across groups?	1 1				v
	4.4 were reasons for withdrawars similar across groups?	4.4	1	1		Λ

	If there is a large attrition from one group and not others, you would want to look for a reason why: the answer to this auestion would then be no					
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5				Х
	test under study?					
	The less under study should be computed to reference less all the time. An example of this might be using a DFX4 machine to measure percent body fat only if a subject's					
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30%$					
5.	Was blinding used to prevent introduction of bias?		Y	N	UN	N
	·		S	U	C	А
					E	
			Х		R	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1	Х			
	to treatment group, <u>as appropriate</u> ?					
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
	studied the effect of MINT on lipid levels in hypercholesterolemic patients. It was					
	and who was not Therefore you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2	Х			
	using an objective test, such as a lab value, this criterion is assumed to be met.)					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	<i>collecting the data).</i>	5.2				v
	factors blinded?	5.5				Λ
	Answer ves, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	influenced by exposure status?					
	<i>Establish who is a case and who is a control at the beginning of the study.</i>	5.5				v
6	Were intervention/therapeutic regimens/exposure factor or procedure and any	5.5	Y	N	U	Λ N
0.	comparison(s) described in detail? Were intervening factors described?		E S	0	N C	А
	······································				L E	
			Х		A R	
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1	Х			
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2				Х
	described?					ļ
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
	produce a meaningful effect?					
	difference in lab values for cholesterol: however 12 days would not be long					
	enough)					
	6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.4	Х			
	How long did the treatment last? Did the patient follow directions?					
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5	Х			
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)			$\vdash$		
1	6.6 Were extra or unplanned treatments described?	6.6				Х
1	<i>i ne iexi may not describe any unplannea treatments. If yes, it would likely be in the discussion section. It is likely there were no unplanned treatments, so a "no"</i>					
1	answer is not a problem overall					
-	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7		<u> </u>		Х
1	For a study to be valid and unbiased, it is important that this be yes.					-

	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8				Х
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L A R X	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	7.2 Were nutrition measures appropriate to question and outcomes of concern? Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.	7.2				X
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4			Х	
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability</i> .	7.5		Х		
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6		Х		
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A R X	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	Х			
	<ul> <li>8.2 Were correct statistical tests used and assumptions of test not violated?</li> <li>You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).</li> </ul>	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)	8.3	X			
	8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)? Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i> . If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.	8.4		X		
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5		X		

8.6 Was clinical significance as well as statistical significance reported? <i>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was</i> <i>reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical</i> <i>significance (P-value) and clinical significance (compare to standard of &lt; 200</i> <i>mg/do for normal cholesterol). A problem can occur when only statistical</i> <i>significance is reported. Reducing cholesterol from 300 to 250 might be statistically</i> <i>significant, but clinically it is still abnormal.</i>	8.6	X			
<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li>Type II (β error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				Х
9. Are <u>conclusions</u> supported by results with biases and limitations taken into		Y E	N O	U N	N A
consideration?		s		C L	
		Х		E A R	
9.1 Is there a discussion of findings?	9.1	Х			
Answer yes or no.					
9.2 Are biases and study limitations identified and discussed?	9.2	X			
This will be in the discussion of finding section that follows the results		v	N	U	N
10. Is bias due to study's <u>funding or sponsorship</u> unlikely? Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		E S	0	N C L E	A
		Х		A	
<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X		K	
10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.	10.2	X			
SYMBOL					
MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report should minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (a)	be design	ated	with	i a	
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is excern	tionally st	trong	the	,	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Citation	Chum, J., Kim, M., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., Samaan, Z. (2017). Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. Evidence-Based Mental Health, 20(4), 128–133. doi:10.1136/eb-2017-102763			
Study Design	Randomized controlled trial			
Class	Α			
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\boxtimes \otimes$ (Neutral)			
Research Purpose	To understand patients' perceived benefit from the Fitbit One and explore themes associated with patient experiences. To compare perceived benefit, patient factors, Fitbit usage, and Beck's Depression Inventory (BDI) scores.			
Inclusion Criteria	Patients 18 years or older with major depressive disorder were approached to participate in the Behavioural Activation Group Program in Patients with Depression (BRAVE) study. All patients with depressive disorder receiving treatment for depression, including pharmacotherapy and psychotherapy were eligible to participate. Among the 87 participants who completed the BRAVE study, 36 participants completed interviews.			
Exclusion Criteria	Inability to provide written informed consent, inability to understand written and spoken English, and having a primary diagnosis other than depression			
	Recruitment: Patients who were attending or referred for an assessment at			
	the mood disorders outpatient clinic			
	Design: During the BRAVE study, patients were asked to carry a Fitbit			
Description of	One for 18-weeks. Interviews were conducted with 36 patients who			
	completed the BRAVE study. Thematic analyses were conducted on the			
	interviews and exploratory quantitative analyses were conducted on			
	patient characteristics, Fitbit usage, steps recorded, perceived benefit, and			
Study Protocol	BDI scores.			
	Blinding used (if applicable): Research assistants collected, recharged,			
	and synced data from each participants' Fitbit One to their online account.			
	Personalized printouts were shared with participants to minimize data			
	manipulation. Research assistants also conducted individual, anonymized			
	interviews with patients who completed the BRAVE study.			

Intervention (if applicable): Fitbit One activity trackers were provided to all patients allocated to the behavioural activation (BA) group at the beginning of the BRAVE study, and email accounts were created for each participant. The patients were encouraged to carry the Fitbit One at all times throughout the 18 weeks to track their physical activity. At the beginning of each BA session, Fitbits were collected, recharged, and synced to the online accounts of their respective holders. The participants' weekly activity data were then exported from the Fitbit website, and a personalized printout was generated including summaries of step count, stairs climbed, and distance travelled. The printouts were shared with participants as they did not have access to the online accounts to minimize data manipulation. However, the Fitbit One can show the number of steps taken daily, which can be seen by the patient. Qualitative data were collected through individual, semi-structured interviews, which aimed to explore patients' experiences with their BA treatment and their Fitbit use. Additional data such as baseline BDI scores, pretherapy goals, familiarity with technology, and other patient characteristics were collected through questionnaires facilitated by a clinician during the study intake. BDI scores were collected weekly throughout the study including at completion. Exploratory analyses were conducted to examine how perceived benefit related to participants' baseline characteristics, including age, BDI scores, pretherapy physical activity goals, and smartphone use, as well as percent change in BDI scores from baseline to completion and Fitbit usage. Statistical Analysis: The control and intervention groups were examined through chi-square tests. To examine the relationships between baseline characteristics and Fitbit use, independent t-tests were used. Between

Pearson's correlation was used to compare Fitbit usage with age and baseline BDI scores.

group differences were also calculated through independent t-tests.

Data Collection Summary	Timing of Measurements: Trained clinicians delivered BA to participants				
	randomized to receive the intervention up to twice per week over 18				
	weeks, for a total of 28 sessions. Patients were asked to carry a Fitbit One				
	at all times for the 18-week intervention period. Individual interviews				
	lasted 30 to 60 minutes and were finished within two months of				
	participants' completion of the BA program.				
	Dependent Variables: Patients' perceived benefits, Fitbit usage, BDI				
	scores				
	Independent Variables: BA topics, such as value assessment, goal setting,				
	leisure education				
	Control Variables: Receiving treatment for depression as usual, including				
	pharmacotherapy and psychotherapy				
	Initial: 36 (18 Males 18 Females)				
	Attrition (final N): 36				
Description of Actual Data Sample	Age: Mean age: 53 years				
	Ethnicity: Unclear				
	Other relevant demographics: Mean BDI score: 36.27 (scale of 0 to 63),				
	23 participants had prior smartphone use, 26 participants attended a				
	college, university, or higher education				
	Anthropometrics: Mean BMI: 30.16 kg/m2				
	Location: Mood disorders outpatient clinic in Hamilton, Ontario, Canada				
	Key Findings: 23 patients found the Fitbit One to be helpful for their				
	physical activity. Themes of positive experiences included 27 patients				
	who reported self-awareness, three patients who reported peer motivation,				
	five patients who reported goal-setting opportunities, and 13 patients who				
Summary of	reported enjoying using the Fitbit One. Themes of negative experiences				
Results	included 12 patients who reported inconvenience, 11 patients who				
	reported inaccuracies, two patients who reported discouragement, and 10				
	patients who reported disinterest. No significant differences were found				
	in age (p=0.72), baseline BDI scores (p=0.44), percent change in BDI				

	scores (p=0.283), pretherapy physical activity goals (p=0.549), and
	smartphone use (p=0.825) between those who did and did not find the
	Fitbit One helpful. However, there was a significant relationship between
	total Fitbit One usage and perceived benefit. The mean number of weeks
	of Fitbit One use for those who found the Fitbit helpful was 18.57 (SD
	1.21) and 12.27 (SD 5.76) weeks for those who did not (p<0.001).
	Other Findings:
	Of the 36 patients who underwent the BRAVE study and completed
Author	interviews, the majority (64%) found the Fitbit One to be helpful,
	identifying their Fitbit use as a factor in improving their physical activity
	level. Many positive themes were concordant with current literature;
	however, patients also reported negative aspects that may affect use.
	Interestingly, there was no significant relationship between perceived
Conclusion	benefit of the Fitbit and percent change in BDI scores, which contradicts
	previous literature supporting Fitbit use in treating depression. These
	findings suggest that the Fitbit One may be useful for patients with
	varying characteristics. Clinicians and researchers should consider both
	strengths and limitations of activity trackers when implementing them to
	motivate patients with depression.
	Strengths: zero percent attrition, use of baseline data and final interviews
	for detailed participant feedback, and use of Fitbit data to corroborate
	adherence
Reviewer	Limitations: small sample size compared to the number of participants
Comments	
Comments	who completed the BRAVE study, difficulty interpreting exploratory
Comments	who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness and
Comments	who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness and there were several confounders such as combined BA intervention, study
Comments	who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness and there were several confounders such as combined BA intervention, study setting, and participants' restricted use of Fitbit
Comments	who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness and there were several confounders such as combined BA intervention, study setting, and participants' restricted use of Fitbit Supported by Canadian Institutes of Health Research (CIHR)
Comments Funding Source	who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness and there were several confounders such as combined BA intervention, study setting, and participants' restricted use of Fitbit Supported by Canadian Institutes of Health Research (CIHR) (Randomised Controlled Trials: Mentoring, code number 201303MTP-
(2013–2015) is to support research time. The authors declared no	
------------------------------------------------------------------	
competing interests.	

#### **Quality Criteria Checklist: Primary Research**

Symbols Used	Explanation
1	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
Q	exceptionally week

Select a rating from the drop-down menu  $\checkmark$ 

Re	levance Questions		
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Yes
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

# If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

٧٤	lidity Questions		
1.	Was the <u>research question</u> clearly stated?	1	Yes
	1.1. Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	Yes
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3. Were the target population and setting specified?	1.3	Yes
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Yes
	2.1. Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes
	without omitting criteria critical to the study? 2.2. Were criteria applied equally to all study groups?	2.2	Yes
	<ul> <li>2.3. Were health, demographics, and other characteristics of subjects described?</li> <li>2.4. Were the subjects/patients a representative sample of the relevant</li> </ul>	2.3	Yes
	population?	2.4	Unclear
3.	Were <u>study groups comparable</u> ? 3.1. Was the method of assigning subjects/patients to groups described and	3	Unclear
	unbiased? (Method of randomization identified if RCT) 3.2. Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	Unclear
	action programmar across study proups at pascine :	3.2	Unclear

rical 3.3	No
on important	
ed for by using 3.4	N/A
ble for cases 3.5	5 N/A
cross-	
2.0	
han 3.6	N/A
4	N/A
st to follow 4.1	N/A
escribed for	
4.2	N/A
nted for? 4.3	8 No
4.4	N/A
indent on	
4.5	N/A
investigators 5	Yes
Investigators	
ne is	Yes
n is assumed	Vac
	1 65
utcomes and 5.3	N/A
inment not	
5.4	N/A
d other test	
5.5	N/A
and any 6	Yes
ragimons (1	Var
regimens 6.1	Yes
6.2	2 N/A
63	Ves
ctor sufficient	105
6.4	Yes
6.5	Yes
described?	
6.6	N/A
	3.7.1.1
or all groups? 6.7	/ N/A
tion 6.7	V N/A
tion 6.8	N/A N/A
for all groups?     6.7       tion     6.8 <b>2?</b> 7	<ul> <li>N/A</li> <li>N/A</li> <li>Unclear</li> </ul>
for all groups?         6.7           tion         6.8           e?         7           the         7.1	<ul> <li>N/A</li> <li>N/A</li> <li>Unclear</li> <li>Yes</li> </ul>
	rical 3.3 on important ed for by using 3.4 ble for cases ontrol, this e cross- h an 4 st to follow 4.1 escribed for 4.2 nted for? 4.3 endent on 4.5 nvestigators 5 ne is 15 ne is 5.1 nue is 5.1 nue is 5.1 nue is 5.2 utcomes and 5.2 utcomes and 5.3 inment not 5.4 d other test 5.5 and any 6 regimens 6.1 ctor sufficient 6.4 pliance 6.5

7.3.	Was the period of follow-up long enough for important outcome(s) to occur?	7.3	N/A
7.4.	Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?	7.4	Unclear
7.5.	Was the measurement of effect at an appropriate level of precision?	7.5	No
7.6.	Were other factors accounted for (measured) that could affect outcomes?	7.6	No
1.1.	7. Were the measurements conducted consistently across groups?	7.7	N/A

8.	8. Was the <u>statistical analysis</u> appropriate for the study design and type of outcome		Unclear
	8.1. Were statistical analyses adequately described the results reported		Yes
	appropriately? 8.2. Were correct statistical tests used and assumptions of test not violated?	8.2	Yes
	8.3. Were statistics reported with levels of significance and/or confidence interv 8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was the	als? 8.3	Yes
	an analysis of outcomes for those maximally exposed or a dose-response	8.4	No
	8.5. Were adequate adjustments made for effects of confounding factors that	8.5	No
	might have affected the outcomes (e.g., multivariate analyses)? 8.6. Was clinical significance as well as statistical significance reported?	8.6	Yes
	8.7. If negative findings, was a power calculation reported to address type 2 error?	or? 8.7	N/A
9.	Are conclusions supported by results with biases and limitations taken into	9	Yes
	consideration?	9.1	Yes
	9.1. Is there a discussion of findings?	92	Ves
10	9.2. Are blases and study limitations identified and discussed?	10	Vez
10.	10.1 Ware sources of funding and investigatory officiations described?	10	Yes
	10.1. Were sources of funding and investigators affiliations described?	10.1	Yes
	10.2. was there no apparent connict of interest?	10.2	Yes

#### MINUS/NEGATIVE (-)

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

#### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

## **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Cook, J., Prairie, M., & Plante, D. (2017). Utility of the Fitbit Flex to Evaluate Sleep in Major Depressive Disorder: A comparison against polysomnography and wrist-worn actigraphy. <i>Journal of Affective Disorders</i> , 217, 299–305. doi:
	10.1016/i.iad.2017.04.030
<b>Study design:</b> Use algorithm – RCT, cohort, etc	Validity study
Study Class (A.B.C.D)	С
Research Quality Rating	POSITIVE (+)
This rating tells if the research	
design is good (+), had (-) or	
neutral (Ø)	
<i>This is determined by the quality</i>	
criteria list. Delete the ratings	
that do not apply (i.e. if positive,	
delete minus/negative and	
neutral).	
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the	To evaluate the utility of the Fitbit Flex (FBF) to estimate sleep in a well-
research question being	characterized cohort of adult patients with major depressive disorder (MDD)
investigated in the study?)	relative to gold standard polysomnography (PSG) and validated actigraphy
	(Actiwatch-2; AW-2).
Inclusion criteria: requirements	Right-handed, unmedicated patients with unipolar MDD who completed an initial
for study eligibility	phone screening, in-person medical, sleep, and psychiatric evaluation. If patients
	met criteria for other psychiatric disorders, MDD had to be considered the primary
	disorder for study inclusion.
Exclusion criteria (conditions	Smoking of >15 cigarettes per day; >3 caffeinated beverages per day; significant
that make individual ineligible)	sleep, neurologic, or medical disorder; history of significant head trauma or loss of
	consciousness >30 minutes; and imminent risk of self-narm of suicide. Women
	pregnant during the study were excluded. Participants were also excluded if they
	met DSM-IV criteria for alcohol or substance abuse/dependence within the
	preceding 6 months
Recruitment	A convenience sample of patients with MDD was recruited as part of a larger study
	investigating electroencephalographic biomarkers of sleep disturbance in
	neuropsychiatric disorders
Blinding used: some of the	A registered sleep technologist, blind to the FBF and AW-2 staging output, staged
persons involved are prevented	all sleep recordings according to standard criteria according to American Academy
from knowing certain information	of Sleep Medicine criteria
that might lead to conscious or	
unconscious bias on their part,	
invalidating the results	
Description of study protocol	I wenty-one patients with unipolar MDD wore the FBF and AW-2 during in-
Intervention: Describe	After an initial phone correcting participants completed on in parson medical
interventions regimens risk	After an initial phone screening, participants completed an in-person medical,
factors or procedures studied	DSM-IV (SCID) semi-structured sleep disorders evaluation and physical evan
Jucions, or procedures studied.	Participants completed additional self-report instruments including the Reck
	Depression Inventory (BDI-II) Pittsburgh Sleen Quality Index (PSOI) and
	Insomnia Severity Index (ISI). Eligible participants were then scheduled for an in-
	laboratory PSG at least one week but no more than one month after their in-person
	screening visit. All participants were instructed to maintain their usual sleep-wake
	schedules for the duration of their time in the study. Participants arrived at

	approximately 18:00 on the night of their PSG for set-up. A wrist-worn AW-2 and
	FBF were both placed adjacently on the participant's non-dominant (left) wrist.
	Polysomnographic data were collected using an integrated recording system that
	utilized a 256-channel EEG net along with other standard recording sensors
	including electrooculogram (EOG) sub-mental electromyogram (EMG)
	electrocardiogram (ECG), bilateral tibial EMG, respiratory inductance
	nlethysmography pulse ovimetry and a position sensor. A registered sleep
	technologist blind to the EPE and AW 2 staging output staged all sleep
	using 20 second enable seconding to standard criteria seconding to American
	using 30-second epochs according to standard criteria according to American
	Academy of Steep Medicine criteria. Bedumes were tailored to each participant's
	nabitual sleep pattern, with lights-off (participant actively trying to fail asleep)
	occurring between approximately 22:00 and 23:00. Participants were allowed to
	sleep ad libitum, remaining undisturbed throughout the night and not awoken at a
	prescribed time the following morning. Lights-on was determined based on the
	participant's stated desire to terminate the nocturnal sleep period upon awakening.
	Polysomnography and accelerometer data were collected within a local network of
	computers time synchronized to an external clock. The following sleep variables
	were calculated for PSG, FBF, and AW-2: total sleep time (TST), sleep onset
	latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE). AW-2
	data were analyzed utilizing the medium threshold relative to PSG, and FBF data
	were analyzed using both the normal and sensitive settings relative to PSG since
	these settings have been shown to produce the most accurate outputs.
Statistical analysis: <i>List tests</i> ,	Bland-Altman analysis was utilized to calculate the mean difference between
significance level set a priori	devices (AW-2 vs. PSG; FBF-N vs. PSG; FBF-N vs. AW-2; FBF-S vs. PSG; and
( $\alpha$ =0.05; include intent to treat	FBF-S vs. AW-2) for each sleep variable (TST, SOL, WASO, SE). Epoch-by-
analysis if applicable; note if	epoch analysis further evaluated sensitivity, specificity, and accuracy for the FBF
there is Power analysis.	and AW-2 relative to PSG. Alpha equaled 0.05 for statistical significance for all
	comparisons.
Timing of measurements: when	Lights-off occurred between approximately 22:00 and 23:00. All sleep recordings
outcomes were measured; usually	were collected and staged in 30-second epochs according to standard criteria.
baseline and one or more later	Participants were not awoken at a prescribed time the following morning.
times	
Dependent variables: outcomes	Sleep variables according to FBF and AW-2
that are measured or registered.	
man are measured or registered,	
variable whose change or	
variable whose change or different states the researcher	
variable whose change or different states the researcher wants to understand, explain, or	
variable whose change or different states the researcher wants to understand, explain, or predict	
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Participants' anthropometrics or stress levels
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand) Control Variables	Participants' anthropometrics or stress levels Sleep variables according to PSG
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand) Control Variables Examples: 1) multivariate logistic	Participants' anthropometrics or stress levels Sleep variables according to PSG
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand) Control Variables Examples: 1) multivariate logistic regression controlled for age,	Participants' anthropometrics or stress levels         Sleep variables according to PSG
variable whose change or different states the researcher wants to understand, explain, or predict Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand) Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin;	Participants' anthropometrics or stress levels         Sleep variables according to PSG
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin;</li> <li>2) usual care; 3) isocaloric diet,</li> </ul>	Participants' anthropometrics or stress levels         Sleep variables according to PSG
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.</li> </ul>	Participants' anthropometrics or stress levels Sleep variables according to PSG
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.</li> <li>Initial n (e.g. 731 (298 males,</li> </ul>	Participants' anthropometrics or stress levels         Sleep variables according to PSG         21 participants, 4 males and 17 females
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.</li> <li>Initial n (e.g. 731 (298 males, 433 females))</li> </ul>	Participants' anthropometrics or stress levels         Sleep variables according to PSG         21 participants, 4 males and 17 females
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin;</li> <li>2) usual care; 3) isocaloric diet, etc.</li> <li>Initial n (e.g. 731 (298 males, 433 females))</li> <li>Record number that entered</li> </ul>	Participants' anthropometrics or stress levels         Sleep variables according to PSG         21 participants, 4 males and 17 females
<ul> <li>variable whose change or different states the researcher wants to understand, explain, or predict</li> <li>Independent variables (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)</li> <li>Control Variables</li> <li>Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.</li> <li>Initial n (e.g. 731 (298 males, 433 females))</li> <li>Record number that entered study – not the number screened.</li> </ul>	Participants' anthropometrics or stress levels         Sleep variables according to PSG         21 participants, 4 males and 17 females

number of subjects that completed	
study	$M_{\text{cons}} = 26.5 \pm 4.6 \text{ years}$
Age usually mean or range	Mean age $-20.3 \pm 4.0$ years
Other relevant demographies:	$M_{\text{eqn}} \text{ RDL II score} = 22.0 \pm 6.8$
demographics describe the	Mean DSOL score $= 8.4 \pm 2.5$
nonulation (students athletes etc)	Mean ISI score = $143 \pm 56$
Anthronometrics: e.g. were	Unclear
groups same or different on	
important physical measures	
(BMI, fitness level)	
<b>Location:</b> <i>Where did the study</i>	Wisconsin Sleep Center
take place? City or country	
Summary of Results: Abstract	When the AW-2 was compared to PSG, AW-2 significantly overestimated TST
results including quantitative data	(mean difference of 40.6 min, p=0.0004) and SE (mean difference of 7.0%,
and statistics. Include statistical	p=0.0003), while significantly underestimating SOL (mean difference of -13.5
significance: P-values, confidence	min, p=0.012) and WASO (mean difference of -27.1 min, p=0.005). When
intervals (CI), relative risk (RR),	compared epoch-by-epoch against PSG, the AW-2 displayed relatively good
odds ratios (OR), likelihood ratio,	sensitivity $(0.97 \pm 0.02)$ and accuracy $(0.87 \pm 0.06)$ , with poor specificity $(0.31 \pm 0.06)$
number needed to treat, power	0.15). When the FBF-N was compared to PSG, like the AW-2, FBF-N
analysis if available.	significantly overestimated TST (mean difference of 46.0 min, p<0.0001) and SE
	(mean difference of 8.1%, $p<0.0001$ ), while significantly underestimating WASO
	(mean difference of $-44.0 \text{ min}$ , p<0.0001). However, SOL assessed by FBF-N and DSC assessed by FBF-N and
	PSG were quite similar (mean difference of $-2.0 \text{ min}$ , $p=0.72$ ). When compared a most by smooth against PSC, again like the AW 2, the EDE N showed a high
	epoch-by-epoch against PSO, again like the AW-2, the PDF-N showed a high sensitivity $(0.98 \pm 0.02)$ and accuracy $(0.88 \pm 0.05)$ with low specificity $(0.35 \pm 0.05)$
	$(0.33 \pm 0.02)$ and accuracy $(0.38 \pm 0.03)$ , with low specificity $(0.35 \pm 0.13)$ . Direct comparison of the EBE-N to AW-2 demonstrated significantly higher
	estimates of SE (mean difference of 1.1% n=0.042) and SOI (mean difference of
	11.5 min n=0.0003) for the FBF-N as well as significantly lower estimates of
	WASO (mean difference of $-16.9 \text{ min } \text{p} \le 0.0001$ ) FBF-N and AW-2 had
	comparable estimates of TST (mean difference of 5.4 min, $p=0.08$ ). When the
	FBF-S was compared to PSG, findings were quite different from those derived
	using the normal mode for the device. Relative to PSG, FBF-S significantly
	underestimated TST (mean difference of -86.3 min, p<0.0001) and SE (mean
	difference of -16.0%, p<0.0001), while significantly overestimating SOL (mean
	difference of 11.5 min, p=0.012) and WASO (mean difference of 74.8 min,
	p<0.0001). When compared epoch-by-epoch against PSG, the FBF-S displayed a
	modest sensitivity (0.78 $\pm$ 0.09), specificity (0.80 $\pm$ 0.17), and accuracy (0.78 $\pm$
	0.08). Similarly, when the FBF-S was compared to the AW-2, FBF-S had
	significantly lower estimates of TST (mean difference of -126.8 min, p<0.0001)
	and SE (mean difference of $-22.9\%$ , p<0.0001) with significantly higher estimates
	of SOL (mean difference of 24.9 min, p=0.0006) and WASO (mean difference of
	101.9 min, p<0.0001).
Author conclusions, navanhugas	Author's Conclusions
Author conclusion: paraphrase	relative to PSC. In the normal setting, the EPE significantly overestimated sleep
had stated by study duthor in	time and efficiency and displayed low specificity. In the sensitive setting, the EBE
body of the report of dostract	significantly underestimated sleep time and efficiency relative to PSG. The FBF is
	not an adequate substitute for PSG when quantifying sleep in MDD however the
	FBF does demonstrate similar performance characteristics to a standard actigraph
	particularly in the estimation of total sleep duration, when used in the normal mode.
	The capabilities, limitations, and settings of the FBF should be carefully considered
	prior to clinical and research implementation.
Reviewer comments: Note	Strengths: zero percent attrition, use of a variety of initial screening methods,
strengths and limitations of study;	including physical examinations and validated questionnaires, and use of AW-2 as

identify concerns that affect study	an alternate measure for sleep and wake to circumvent some of the shortcomings of					
validity and generalizability—	PSG.					
your comments should be	Limitations: participants were young to middle aged and predominantly female,					
italicized)	which may limit generalizability of findings. Study specifically examined					
	outpatients with MDD, thus findings cannot be directly extended to other mood					
	and/or sleep disorders. Also, results cannot be extended to other fitness trackers, or					
	more current generations of the same model as these devices may have different					
	erformance characteristics. The study design leaves in question the true					
	apabilities of the FBF as a long-term sleep measurement device.					
	Funding source: this work was supported by grants from the National Institute of					
	Mental Health (K23MH099234), the Brain and Behavior Research Foundation, and					
	the American Sleep Medicine Foundation. The sources of funding for this					
	investigation had no further role in the study design, data collection, analysis and					
	interpretation of the data, and the decision to submit the paper for publication. The					
	study authors have no relationship with Fitbit, and Fitbit did not supply any					
	funding, supplies, or guidance towards this investigation, or have any bearing on					
	the decision to submit this manuscript for publication.					

Table 3.2.a. Quality Criteria Checklist: Primary Research

DELEVANCE OUESTIONS					
<b>RELEVANCE QUESTIONS</b>	1	Y	N	U	Ν
Cook I. Prairie M. & Plante D. (2017). Utility of the Fithit Fley to Evaluate Sleep in Major		E	0	N	А
Depressive Disorder: A comparison against polycompography and wrist worn actigraphy		3		Ľ	
Lowned of Affective Disordere 217 200, 205, doi: 10.1016/j.jod.2017.04.020				E A	
<i>Journal of Affective Disorders</i> , 217, 299–303. doi: 10.1010/j.jad.2017.04.030				R	
1. Would implementing the studied intervention or procedure (if found successful) result in	1			X	
improved outcomes for the patients/clients/population group? (Not Applicable for some					
epidemiological studies)	-				
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	Х			
population group would care about?	-				
3. Is the focus of the intervention or procedure (independent variable) or topic of study a	3		Х		
common issue of concern to dietetics practice?					
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	X			
If the answers to all of the above relevance questions are "yes", the report is eligible for design	<i>ation</i> w	vith a	plu	s (+)	on
the Evidence Quality Worksheet, depending on answers to the following validity questions.					
VALIDITY QUESTIONS	-				
1. Was the <u>research question</u> clearly stated?		Y	I	N U D N	
This is usually stated at end of the introduction and just before methods section.		s		Ċ	
				E	
		X		A	
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified?	1 1	v	-		
This is often called the treatment and explained in the methods section	1.1	Δ	•		
1.2 Was the outcome(a) (dependent variable(a)) algority indicated?	1.2	v	-		-
These are sometimes called the and points: the results section reports the outcomes, but	1.2	Δ	•		
this information should be in the methods section too					
1.2 Were the target population and setting specified?	1.2	v	-		_
The target nonulation is group for whom findings may be applicable: look for this in the	1.5	Δ	•		
introduction and in the methods section					
2 Was the selection of study subjects/petients free from bias?		Y	ľ	v u	N
2. Was the <u>selection</u> of study subjects/patients free from blas:		E	0		A
		5		L	
		x	-	A	
2.1. Ware inclusion/analysism oritaria analified (a.g. risk point in diasage programmer)	2.1	v	-	R	<u> </u>
2.1 were inclusion/exclusion cineria specified (e.g., fisk, point in disease progression,	2.1	Δ	-		
aritaria aritiari to the study?					
The authors should give several points about the inclusion/avalusion aritaria such as the					
The duthors should give several points about the inclusion/exclusion criteria such as the					
inclusion Exclusion criteria should be listed too although some are					
understood For example if the ages for inclusion are 18 to 70 the authors will					
probably not specifically note that children and people over age 70 were					
excluded Most of the time however subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD)					
2.2 Were criteria applied equally to all study groups?	2.2				x
2.2 Were health demographics and other characteristics of subjects described?	2.2	X	-		
There is usually a Table 1 summarizing demographics and characteristics at baseline	2.5	1			
Groups are not different if the P-Value is $> 0.05$ If there has been a previous					
paper describing the study population that paper may be referenced and you					
would need to go back to the original nublication to see that Table 1					
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4		+	3	ζ
The abstractor may have to apply a bit of clinical iudgment here. Authors try to be brief and may				1	-
only say that the patients came from the same clinic from people who met the inclusion					
criteria			1		

3. Were study groups comparable?		Y	N O	U N	N A
There is usually a Table 1 summarizing demographics and characteristics at baseline.		s	Ŭ	C	
Groups are <u>not</u> different if the P-Value is $> 0.05$ .				E	
				A R	Х
3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
(Method of randomization identified if RCT)					
In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
randomized block design; or assigned by computer-generated random numbers).					
Look for instances that show bias; for example I once read a study where patients					
were randomized to receive liquid energy supplements; however, if someone					
disliked their supplement, they were allowed to change groups – this is not unbiased	7				
3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
demographics) similar across study groups at baseline? See Table I for this - there					
should be no significant differences across study groups in an intervention study.					
3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Х			
Most RCTs use a concurrent control group. Occasionally an intervention study will use a	1				
prior study as a control group; that is an example of a historical control. That is not					
as strong a research design as use of concurrent control group. A crossover study					
where the subject acts as his/her own control is use of concurrent control.					
3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4				Х
confounding factors and/or were preexisting differences accounted for by using					
appropriate adjustments in statistical analysis?					
The groups in a cohort or cross-sectional study should not be different from each other;					
if they are, a strong study will utilize statistical techniques such as multivariate					
analyses to remove the variance due to the group differences. Look for this					
information in the statistics and results sections.					
3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
controls? If case series or trial with subjects serving as own control, this criterion is					
not applicable. Criterion may not be applicable in some cross-sectional studies.					
Subjects are generally matched for age, gender, etc. Look for this in the statistical					
description and results sections.					
3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	Х			
reference standard (e.g. "gold standard")?					
Example: comparing body fat analysis method with underwater weighing (gold					
standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
Calorimetry) is used.					
4. Was method of handling withdrawals described?		Y	N	U	N
		S	0	C	А
				L E	
				A	Х
4.1 Were follow up methods described and the same for all groups?	4.1			ĸ	X
4.2 Was the number characteristics of withdrawals (i.e. dropouts, lost to follow up	4.2				X
attrition rate) and/or response rate (cross-sectional studies) described for each group	) 7.2				1
(Follow up goal for a strong study is 80 %)	,				
This should be found in the results section. If there is attrition $> 20\%$ it is important to					
note that on the worksheet (as a note in the results section or in the reviewer					
comments at the very hottom)					
4.3 Were all enrolled subjects/natients (in the original sample) accounted for?	43	x			
This information is often presented in a figure with # recruited # enrolled (this is initial	т.5	Λ			
N # remaining at end of study period (final N). Sometimes the reasons that subjects	,				
withdrew or were dronned is given in the figure or in the text (results section)					
4.4 Were reasons for withdrawals similar across groups?	4 4			$\vdash$	x
If there is a large attrition from one group and not others you would want to look for a	7.7				11
reason why: the answer to this question would then be no					
reason may, me answer to this question noura men de no.			1	1 1	

	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	Х			
	test under study? The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30\%$ .					
5.	Was blinding used to prevent introduction of bias?		Y	N	U	N
			S	0	C	А
					L E	
			Х		A R	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1				Х
	to treatment group, as appropriate?					
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
	studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					
	an RCT, but obviously, the subjects and practitioners knew who was getting MNT					
	and who was not. Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the distitians and patients to know they were receiving MNT.	5.2	V			
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an abjective test, such as a lab value, this criterion is assumed to be met.)	5.2	Χ			
	Answer was if a lab test was used to measure an outcome. A method of blinding a dist					
	study is to have senarate neonle analyzing the data (not the same ones who were					
	collecting the data)					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	influenced by exposure status?					
	<i>Establish who is a case and who is a control at the beginning of the study.</i>	5.5			v	
6	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5	Y	N		Ν
0.	comparison(s) described in detail? Were intervening factors described?		E	0	N C	А
	comparison(s) described in detail. Were intervening factors described.		5		L	
			Х		A	
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1			ĸ	Х
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2	Х			
	described?					
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
	produce a meaningful effect?					
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
	difference in lab values for cholesterol; however, 12 days would not be long					
	6.4 Was the amount of exposure and if relevant subject/nations compliance measured?	6.4	v			
	How long did the treatment last? Did the patient follow directions?	0.4	Λ			
	6.5 Were co-interventions (e.g. ancillary treatments other therapies) described?	6.5				X
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)					
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					
	discussion section. It is likely there were no unplanned treatments, so a "no"					
	answer is not a problem overall.	<u> </u>		$\square$		<u> </u>
1	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7				Х
	For a study to be valid and unbiased, it is important that this be yes.		37	$\left  - \right $		
1	0.8 in diagnostic study, were details of test administration and replication sufficient?	0.8	X			
	Osuany answer n/a jor alei suay.	I				ı

7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E	N A
			Х		A R	
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures? Check that surveys were validated	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? Precision is reproducibility or repeatability.	7.5	X			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A	N A
	9.1 Ware statistical analysis of successful described and the nexulta remembed any maintail. 9	0.1			R	
	There should be a discussion of the statistics in the methods section.	0.1	Λ			
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	Х			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)	8.3	X			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4				X
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5				Х
	<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical</li> </ul>	8.6	X			

	significance is reported. Reducing cholesterol from 300 to 250 might be statistically significant, but clinically it is still abnormal.					
	8.7 If negative findings, was a power calculation reported to address type 2 error? <i>Type II (<math>\beta</math> error is a false negative that happens when the investigators fail to reject the</i> <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."	8.7				Х
9.	Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S	N O	U N C L E A	N A
	9.1 Is there a discussion of findings?	9.1	X		к	
	9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	X			
10.	<b>Is bias due to study's <u>funding or sponsorship</u> unlikely?</b> Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
	<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X			
	10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.	10.2	X			
SYI	MBOL					
MII If m min NE	MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report should be designated with a minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (0)					

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.



Citation	Cook, J., Prairie, M., & Plante, D. (2017). Utility of the Fitbit Flex to Evaluate Sleep in Major Depressive Disorder: A comparison against polysomnography and wrist-worn actigraphy. Journal of Affective Disorders, 217, 299–305. doi: 10.1016/j.jad.2017.04.030				
Study Design	Validity study				
Class	С				
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\otimes$ (Neutral)				
Research Purpose	To evaluate the utility of the Fitbit Flex (FBF) to estimate sleep in a well- characterized cohort of adult patients with major depressive disorder (MDD) relative to gold standard polysomnography (PSG) and validated actigraphy (Actiwatch-2; AW-2).				
Inclusion Criteria	Right-handed, unmedicated patients with unipolar MDD who completed an initial phone screening, in-person medical, sleep, and psychiatric evaluation. If patients met criteria for other psychiatric disorders, MDD had to be considered the primary disorder for study inclusion.				
Exclusion Criteria	Smoking of >15 cigarettes per day; >3 caffeinated beverages per day; significant sleep, neurologic, or medical disorder; history of significant head trauma or loss of consciousness >30 minutes; and imminent risk of self-harm or suicide. Women who were pregnant, breastfeeding, <6 months post-partum, or planning to become pregnant during the study were excluded. Participants were also excluded if they met DSM-IV criteria for alcohol or substance abuse/dependence within the preceding 6 months.				
Description of Study Protocol	Recruitment: A convenience sample of patients with MDD was recruited as part of a larger study investigating electroencephalographic biomarkers of sleep disturbance in neuropsychiatric disorders Design: Twenty-one patients with unipolar MDD wore the FBF and AW- 2 during in-laboratory PSG Blinding used (if applicable): A registered sleep technologist, blind to the FBF and AW-2 staging output, staged all sleep recordings according to standard criteria according to American Academy of Sleep Medicine criteria Intervention (if applicable): After an initial phone screening, participants completed an in-person medical, sleep, and psychiatric evaluation that included the Structured Clinical Interview for DSM-IV (SCID), semi-				

structured sleep disorders evaluation, and physical exam. Participants completed additional self-report instruments including the Beck Depression Inventory (BDI-II), Pittsburgh Sleep Quality Index (PSQI), and Insomnia Severity Index (ISI). Eligible participants were then scheduled for an in-laboratory PSG at least one week but no more than one month after their in-person screening visit. All participants were instructed to maintain their usual sleep-wake schedules for the duration of their time in the study. Participants arrived at approximately 18:00 on the night of their PSG for set-up. A wrist-worn AW-2 and FBF were both placed adjacently on the participant's non-dominant (left) wrist. Polysomnographic data were collected using an integrated recording system that utilized a 256-channel EEG net along with other standard recording sensors including electrooculogram (EOG), sub-mental electromyogram (EMG), electrocardiogram (ECG), bilateral tibial EMG, respiratory inductance plethysmography, pulse oximetry, and a position sensor. A registered sleep technologist, blind to the FBF and AW-2 staging output, staged all sleep recordings using 30-second epochs according to standard criteria according to American Academy of Sleep Medicine criteria. Bedtimes were tailored to each participant's habitual sleep pattern, with lights-off (participant actively trying to fall asleep) occurring between approximately 22:00 and 23:00. Participants were allowed to sleep ad libitum, remaining undisturbed throughout the night and not awoken at a prescribed time the following morning. Lights-on was determined based on the participant's stated desire to terminate the nocturnal sleep period upon awakening. Polysomnography and accelerometer data were collected within a local network of computers time synchronized to an external clock. The following sleep variables were calculated for PSG, FBF, and AW-2: total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE). AW-2 data were analyzed utilizing the medium threshold relative to PSG, and FBF data were analyzed using both the normal and sensitive

	settings relative to PSG since these settings have been shown to produce					
	the most accurate outputs.					
	Statistical Analysis: Bland-Altman analysis was utilized to calculate the					
	mean difference between devices (AW-2 vs. PSG; FBF-N vs. PSG; FBF-					
N vs. AW-2; FBF-S vs. PSG; and FBF-S vs. AW-2) for each s						
	variable (TST, SOL, WASO, SE). Epoch-by-epoch analysis further					
	evaluated sensitivity, specificity, and accuracy for the FBF and AW-2					
	relative to PSG. Alpha equaled 0.05 for statistical significance for all					
	comparisons.					
	Timing of Measurements: Lights-off occurred between approximately					
	22:00 and 23:00. All sleep recordings were collected and staged in 30-					
Data Collection	second epochs according to standard criteria. Participants were not					
Summary	awoken at a prescribed time the following morning.					
	Dependent Variables: Sleep variables according to FBF and AW-2					
	Independent Variables: Participants' anthropometrics or stress levels					
	Control Variables: Sleep variables according to PSG					
	Initial: 21 (4 Males 17 Females)					
	Attrition (final N): 21					
	Age: Mean age = $26.5 \pm 4.6$ years					
Description of	Ethnicity: Unclear					
Actual Data Sample	Other relevant demographics: Mean BDI-II score = $22.9 \pm 6.8$ , mean					
Sumpre	PSQI score = $8.4 \pm 2.5$ , mean ISI score = $14.3 \pm 5.6$					
	Anthropometrics: Unclear					
	Location: Wisconsin Sleep Center					
	Key Findings: When the AW-2 was compared to PSG, AW-2					
	significantly overestimated TST (mean difference of 40.6 min, p=0.0004)					
Summary of	and SE (mean difference of 7.0%, p=0.0003), while significantly					
Results	underestimating SOL (mean difference of -13.5 min, p=0.012) and					
	WASO (mean difference of $-27.1$ min, p=0.005). When compared					
	epoch-by-epoch against PSG, the AW-2 displayed relatively good					

sensitivity $(0.97 \pm 0.02)$ and accuracy $(0.87 \pm 0.06)$ , with poor specificity
$(0.31 \pm 0.15)$ . When the FBF-N was compared to PSG, like the AW-2,
FBF-N significantly overestimated TST (mean difference of 46.0 min,
p<0.0001) and SE (mean difference of 8.1%, p<0.0001), while
significantly underestimating WASO (mean difference of -44.0 min,
p<0.0001). However, SOL assessed by FBF-N and PSG were quite
similar (mean difference of -2.0 min, p=0.72). When compared epoch-
by-epoch against PSG, again like the AW-2, the FBF-N showed a high
sensitivity (0.98 $\pm$ 0.02) and accuracy (0.88 $\pm$ 0.05), with low specificity
$(0.35 \pm 0.13)$ . Direct comparison of the FBF-N to AW-2 demonstrated
significantly higher estimates of SE (mean difference of 1.1%, p=0.042)
and SOL (mean difference of 11.5 min, p=0.0003) for the FBF-N, as well
as significantly lower estimates of WASO (mean difference of -16.9 min,
p<0.0001). FBF-N and AW-2 had comparable estimates of TST (mean
difference of 5.4 min, p=0.08). When the FBF-S was compared to PSG,
findings were quite different from those derived using the normal mode
for the device. Relative to PSG, FBF-S significantly underestimated TST
(mean difference of $-86.3$ min, p<0.0001) and SE (mean difference of
-16.0%, p<0.0001), while significantly overestimating SOL (mean
difference of 11.5 min, p=0.012) and WASO (mean difference of 74.8
min,
p<0.0001). When compared epoch-by-epoch against PSG, the FBF-S
displayed a modest sensitivity (0.78 $\pm$ 0.09), specificity (0.80 $\pm$ 0.17), and
accuracy (0.78 $\pm$ 0.08). Similarly, when the FBF-S was compared to the
AW-2, FBF-S had significantly lower estimates of TST (mean difference
of -126.8 min, p<0.0001) and SE (mean difference of -22.9%, p<0.0001)

with significantly higher estimates of SOL (mean difference of 24.9 min,

p=0.0006) and WASO (mean difference of 101.9 min, p<0.0001).

Other Findings:

	The FBF demonstrated significant limitations in quantifying sleep and
	wake, relative to PSG. In the normal setting, the FBF significantly
	overestimated sleep time and efficiency, and displayed low specificity. In
	the sensitive setting, the FBF significantly underestimated sleep time and
Author	efficiency relative to PSG. The FBF is not an adequate substitute for PSG
Conclusion	when quantifying sleep in MDD, however, the FBF does demonstrate
	similar performance characteristics to a standard actigraph, particularly in
	the estimation of total sleep duration, when used in the normal mode. The
	capabilities, limitations, and settings of the FBF should be carefully
	considered prior to clinical and research implementation.
-	Strengths: zero percent attrition, use of a variety of initial screening
	methods, including physical examinations and validated questionnaires,
	and use of $AW$ -2 as an alternate measure for sleep and wake to
	circumvent some of the shortcomings of PSG.
	Limitations: participants were young to middle aged and predominantly
Reviewer	female, which may limit generalizability of findings. Study specifically
Comments	examined outpatients with MDD, thus findings cannot be directly
	extended to other mood and/or sleep disorders. Also, results cannot be
	extended to other fitness trackers, or more current generations of the
	same model as these devices may have different performance
	characteristics. The study design leaves in question the true capabilities
	of the FBF as a long-term sleep measurement device.
	This work was supported by grants from the National Institute of Mental
	Health (K23MH099234), the Brain and Behavior Research Foundation,
	and the American Sleep Medicine Foundation. The sources of funding for
	this investigation had no further role in the study design, data collection,
Funding Source	analysis and interpretation of the data, and the decision to submit the
	paper for publication. The study authors have no relationship with Fitbit,
	and Fitbit did not supply any funding, supplies, or guidance towards this
	investigation, or have any bearing on the decision to submit this
	manuscript for publication.

#### **Quality Criteria Checklist: Primary Research**

Symbols Used	Explanation
1	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
	exceptionally week

Select a rating from the drop-down menu  $\checkmark$ 

Re	Relevance Questions							
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear					
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes					
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No					
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes					
If wi qu	the answers to all of the above relevance questions are "Yes," the report is th a plus (+) on the Evidence Quality Worksheet, depending on answers to estions.	eligil the f	ble for designation following validity					
V٤	lidity Questions							
1.	Was the <u>research question</u> clearly stated?	1	Yes					
	1.1. Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes					
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes					
	1.3. Were the target population and setting specified?	1.3	Yes					
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Yes					

2.1. Were incl		Were inclusion/exclusion criteria specified (e.g., risk, point in disease		
		progression, diagnostic or prognosis criteria), and with sufficient detail and		Yes
	2.2	without omitting criteria critical to the study?	2.2	N/A
	۷.۷.	were criteria applied equally to all study groups?		1011
	2.3.	Were health, demographics, and other characteristics of subjects described?	2.3	Yes
	2.4.	Were the subjects/patients a representative sample of the relevant		
		population?	2.4	Unclear
3.	Were g	study groups comparable?	2	NT/A
	3.1.	Was the method of assigning subjects/patients to groups described and	3	N/A
		unbiased? (Method of randomization identified if RCT)		
	3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.1	N/A

demographics) similar across study groups at baseline? 3.3. Were concurrent controls used? (Concurrent preferred over historical controls.) 3.2

3.4. If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using appropriate adjustments in statistical analysis?
 3.4

N/A

Yes

N/A

		3.5. If case control study, were potential confounding factors comparable for cases and controls? (If case series or trial with subjects serving as own control, this	3.5	N/A
		<ul> <li>criterion is not applicable. Criterion may not be applicable in some cross-sectional studies.)</li> <li>3.6. If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?</li> </ul>	3.6	Yes
Ī	4.	Was method of handling <u>withdrawals</u> described?	4	N/A
		4.1. Were follow up methods described and the same for all groups?	41	N/A
		4.2. Was the number, characteristics of withdrawais (i.e., dropouts, lost to follow up, attrition rate) and/or response rate (cross-sectional studies) described for		
		each group? (Follow up goal for a strong study is 80%.)	4.2	N/A
		4.3. Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Yes
		4.4. Were reasons for withdrawals similar across groups	4.4	N/A
		4.5. If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4.5	Vec
	-	Weshinding used to provent introduction of bios?	ч.5	105
	5.	5.1. In intervention study, were subjects, clinicians/practitioners, and investigators	5	Yes
		5.2. Were data collectors blinded for outcomes assessment? (If outcome is	5.1	N/A
		to be met.)	5.2	Yes
		5.3. In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	5.3	N/A
		5.4. In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4	N/A
		5.5. In diagnostic study, were test results blinded to patient history and other test results?	5.5	Unclear
	6.	Were intervention/therapeutic regimens/exposure factor or procedure and any	6	Yes
		comparison(s) described in detail? Were <u>intervening factors</u> described?	6.1	NI/A
		studied?	0.1	IN/A
		6.2. In observational study, were interventions, study settings, and	6.2	Yes
		clinicians/provider described? 6.3. Was the intensity and duration of the intervention or exposure factor sufficient	6.3	Yes
		to produce a meaningful effect?	6.4	Yes
		measured?	6.5	N/A
		<ul><li>6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described?</li><li>6.6. Were extra or unplanned treatments described?</li></ul>	6.6	N/A
		6.7. Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?	6.7	N/A
		sufficient?	6.8	Yes
F	7.	Were <u>outcomes</u> clearly defined and the <u>measurements valid and reliable</u> ?	7	Yes
		7.1. Were primary and secondary endpoints described and relevant to the question?	7.1	Yes
		7.2. Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A
		7.3. Was the period of follow-up long enough for important outcome(s) to occur?	7.3	N/A
		7.4. Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?	7.4	Yes
ļ		7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
		7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	Yes

lucted consistently across groups? 7.7 N/A	7.7. Were the measurements conduc
Ite for the study design and type of outcome 8 Yes	Was the <u>statistical analysis</u> appropriate indicators?
uately described the results reported 8.1 Yes	8.1. Were statistical analyses adequa
used and assumptions of test not violated? 8.2 Yes	appropriately? 8.2. Were correct statistical tests use
evels of significance and/or confidence intervals? 8.3 Yes	8.3. Were statistics reported with lev 8.4 Was "intent to treat" analysis of
ose maximally exposed or a dose-response 8.4 N/A	an analysis of outcomes for those analysis)?
nade for effects of confounding factors that 8.5 N/A	8.5. Were adequate adjustments ma
mes (e.g., multivariate analyses)? Il as statistical significance reported? 8.6 Yes	might have affected the outcome 8.6. Was clinical significance as well a
ver calculation reported to address type 2 error? 8.7 N/A	8.7. If negative findings, was a power
with biases and limitations taken into 9 Yes	Are conclusions supported by results w
91 Ves	consideration?
s?	9.1. Is there a discussion of findings?
ns identified and discussed? 9.2 Yes	9.2. Are biases and study limitations
isorship unlikely? 10 Yes	0. Is bias due to study's <u>funding or sponso</u>
nvestigators' affiliations described? 10.1 Yes	10.1. Were sources of funding and inv
t of interest? 10.2 Yes	10.2. Was there no apparent conflict of
s?9.1Yesns identified and discussed?9.2Yessorship unlikely?10Yesnvestigators' affiliations described?10.1Yesit of interest?10.2Yesabove validity questions are "No," the report should be designated with a	<ul> <li>9.1. Is there a discussion of findings?</li> <li>9.2. Are biases and study limitations</li> <li><b>D.</b> Is bias due to study's <u>funding or sponso</u> 10.1. Were sources of funding and inv 10.2. Was there no apparent conflict of <b>IINUS/NEGATIVE (-)</b> <i>fimost (six or more) of the answers to the ab</i></li> </ul>

#### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

## **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Ferguson, T., Rowlands, A., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. <i>International Journal of Behavioral Nutrition and Physical Activity</i> , <i>12</i> , 42. doi:10.1186/s12966-015-0201-9
<b>Study design:</b> Use algorithm – <i>RCT, cohort, etc</i>	Validity study
Study Class (A,B,C,D)	С
Research Quality Rating	POSITIVE (+)
This rating tells if the research	
design is good (+), bad (-) or	
neutral (Ø)	
This is determined by the quality	
that do not apply (i.e. if positive	
delete minus/negative and	
neutral).	
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the	To assess the concurrent validity of a selection of consumer-level accelerometer-
research question being	based activity monitors compared to two research-grade accelerometers in free
investigated in the study?)	living conditions
Inclusion criteria: requirements	Healthy adult participants over 18 years of age, living in metropolitan Adelaide,
for study eligibility	South Australia, and could ambulate without walking aids
<b>Exclusion criteria</b> (conditions	Self-reported injury or illness affecting mobility
Recruitment	Unclear, convenience sample
Blinding used: some of the	NA
persons involved are prevented	
from knowing certain information	
that might lead to conscious or	
unconscious bias on their part,	
invalidating the results	21 haalthu adult nortiginanta wara gavan gangumar laval gativitu manitara (Fithit
What happened in the study?	One Fithit Zin Jawhone IIP Misfit Shine Nike Fuelband Strijy Smart Pedometer
mai nappenea in the study.	and Withings Pulse) and two research-grade accelerometers/multi-sensor devices
	(BodyMedia SenseWear, and ActiGraph GT3X+) simultaneously for 48-hours.
	Participants went about their daily life in free-living conditions during data
	collection. Four physical activity parameters were measured, including step count,
	moderate to vigorous physical activity (MVPA), sleep, and total daily energy
Intervention Describe	expenditure (IDEE).
interventions, regimens, risk	Shine Nike Fuelband Strijv Smart Pedometer and Withings Pulse) were examined
factors, or procedures studied.	and chosen based on availability to the authors for purchase between February and
	August 2013. The consumer-level devices were compared with two research grade
	tri-axial accelerometers/multi-sensor devices (BodyMedia SenseWear Model MF
	and ActiGraph GT3X+), which have accepted reliability and validity as free-living
	measures of physical activity and sleep time. Participants attended an appointment
	at which demographic data (date of birth, sex, and dominant side) were obtained, with height and mass measured following standardized procedures. All pine
	devices were fitted to the participant in the following locations: SenseWear on the
	left upper arm; Fuelband, UP and Shine on the left wrist; GT3X+, One, Zip, Pulse
	and Striiv on the right side of the waist on an elasticized belt. Where consumer-
	level devices were designed for multiple wear locations, devices suitable for wrist

	wear were worn on the wrist; otherwise the device was worn on the waist.
	Placement order of the devices at the wrist and waist was randomized. Participants
	were instructed to leave all devices on simultaneously for approximately 48 hours
	(including sleep, but excluding showering) in order to capture a full overnight sleep
	episode as well as two 24-hours of activity data from midnight to midnight. The
	wear period was not limited to a particular period of the week (i.e. not restricted to
	weekdays only or weekends only), and no guidelines or restrictions on activity
	levels or sleep were provided, in order to ensure the study broadly represented free-
	living conditions. Participants were instructed how to turn sleep mode on and off
	for the relevant devices (Shine, Pulse, One, UP). Participants were not given
	access to any of the device software or account information and were also
	instructed not to turn off, modify, or change any device wear locations once fitted.
	Devices were collected after the 48-hour wear period for data collection. Data
	were extracted using the proprietary software for all consumer devices, in the same
	fashion that a consumer would utilize the software.
Statistical analysis: <i>List tests</i> ,	Participants' demographic data were analyzed descriptively. The validity of the
significance level set a priori	consumer-level activity monitors relative to the research-grade devices for step
$(\alpha=0.05; include intent to treat$	count, MVPA, sleep, and TDEE was quantified using Bland-Altman analysis,
analysis if applicable; note if	median absolute difference, and Pearson's correlation. A priori power analyses
there is Power analysis.	were undertaken based on existing data on correlations among various research
	devices, which suggested that the correlation between consumer-level and research
	devices would be about 0.85. If the actual population correlation between
	consumer-level and research-grade devices was 0.85, then a target sample size of 21 would yield in 0.5% of eases, a sample correlation between 0.65 and 0.04
Timing of moogunomontor where	21 would yield, in 95% of cases, a sample correlation between 0.05 and 0.94.
autoomas ware magnined: usually	daviage. However, data relating to physical activity wars limited to the full
baseline and one or more later	calendar day (24 hour period midnight to midnight) following initialization and
times	data relating to sleep were limited to the first night of sleep (24 hour period midday
times	to midday, excluding nans) following initialization
Dependent variables: outcomes	Physical activity parameters (step count_MVPA_sleep_and TDEE) according to
that are measured or registered:	Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike Fuelband, Strijv Smart
variable whose change or	Pedometer, and Withings Pulse
different states the researcher	, , ,
wants to understand, explain, or	
predict	
Independent variables	Free living conditions (e.g. participants' daily obligations, lifestyles, level of
(intervention or procedure; this	physical fitness, stress levels)
variable can be manipulated; a	
variable whose effect upon the	
dependent variable one is trying	
to understand)	
Control Variables	Physical activity parameters (step count, MVPA, sleep, and TDEE) according to
Examples: 1) multivariate logistic	BodyMedia SenseWear and ActiGraph G13X+
regression controlled for age,	
BMI, albumin;	
2) usual care; 3) isocaloric diet,	
etc.	21 maticipants 10 malas and 11 famalas
<b>Initial n</b> (e.g. $731$ (298 males,	21 participants, 10 males and 11 females
Record number that entered	
fully = not the number screened	
<b>Final n</b> (attrition)	21 participants 10 males and 11 females
number of subjects that completed	- · participatio, to make what it females
study	

Ethnicity (if given)	Unclear
Other relevant demographics:	All participants were right hand dominant
demographics describe the	
population (students, athletes, etc)	
Anthropometrics: e.g. were	Male BMI: $27.3 \pm 3.2 \text{ kg/m2}$
groups same or different on	Female BMI: $25.5 \pm 5.2 \text{ kg/m2}$
important physical measures	
(BMI, fitness level)	
Location: Where did the study	Metropolitan Adelaide, South Australia
take place? City or country	
Summary of Results: Abstract	All consumer-level activity monitors measured steps, and correlations with
results including quantitative data	reference devices were very strong ( $r = 0.94-0.99$ ). Bland-Altman analyses
and statistics. Include statistical	suggested that three of the activity monitors slightly over-counted (Striiv, Zip, One)
significance: P-values, confidence	while four under-counted (Fuelband, Shine, Up, Pulse). Five of the activity
intervals (CI), relative risk (RR),	monitors (Striiv, Shine, Up, Zip, One) were considered to measure a parameter
odds ratios (OR), likelihood ratio,	similar or equivalent to MVPA time. Correlations between readings from the
number needed to treat, power	activity monitors and reference devices ranged from weak to strong ( $r = 0.52-0.91$ ).
analysis if available.	Bland-Altman analyses showed large differences between the mean values
	reported. For example, the Shine under-counted (mean = 53.3 min of MVPA
	compared to reference device (GT3X+) mean = 58.5 min), while the Striiv over-
	counted (mean = 249 min of MVPA compared to reference device (GT3X+)). Of
	the five activity monitors (Shine, Up, Pulse, Zip, One) that measured TDEE,
	correlations with the reference devices were moderate to strong ( $r = 0.74 - 0.81$ ).
	Bland-Altman analyses suggest all activity monitors considerably underestimated
	TDEE compared to the reference device (SenseWear, mean = 3005 kcal), ranging
	from 475 kcal (One) to 898 kcal (UP). Of the four activity monitors (Shine, Up,
	Pulse, One) that measured minutes of sleep, all correlated strongly with the
	reference device ( $r = 0.82-0.92$ ). Bland-Altman analyses showed all activity
	monitors overestimated minutes of sleep, most notably, the Shine (mean = 44 min)
	compared to reference device (SenseWear) mean = 424 min).
	Author's Conclusions
Author conclusion: paraphrase	In free-living conditions, the consumer-level activity monitors showed strong
that stated by study author in	validity for the measurement of steps and sleep duration, and moderate-to-strong
body of the report or abstract	valid for measurement of TDEE and MVPA. Median absolute differences were
	generally modest for sleep and steps, moderate for TDEE, and large for MVPA.
	Validity for each construct ranged widely between activity monitors, with the Fitbit
	One, Fitbit Zip, and Withings Pulse being the strongest performers.
Reviewer comments: Note	Strengths: zero percent attrition, the use of numerous consumer and reference
strengths and limitations of study;	devices, testing the devices in free-living conditions as they are designed for, and
identify concerns that affect study	examining several different physical activity variables collected by the devices
validity and generalizability—	Limitations: participant recruitment was not discussed, blinding was not utilized,
your comments should be	and validity may vary if activity monitors are worn in locations other than the hip
italicized)	or wrist
· · · · · · · · · · · · · · · · · · ·	Funding source: unclear, authors declared no competing interests

#### Table 3.2.a. Quality Criteria Checklist: Primary Research

DELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet) Ferguson T. Rowlands A. Olds T. & Maher C. (2015). The validity of consumer-level		Y E S	N O	U N C	N A
activity monitors in healthy adults worn in free-living conditions: a cross-sectional study		5		L F	
International Journal of Behavioral Nutrition and Physical Activity. 12, 42.				A	
doi:10.1186/s12966-015-0201-9				R	
1. Would implementing the studied intervention or procedure (if found successful) result in	1			Х	
improved outcomes for the patients/clients/population group? (Not Applicable for some					
epidemiological studies)					
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	Х			
3 Is the focus of the intervention or procedure (independent variable) or topic of study a	3		v		
common issue of concern to dietetics practice?	5		Λ		
4 Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	X			
If the answers to all of the above relevance questions are "ves", the report is eligible for design	nation w	ith a	plus	(+)	on
the Evidence Quality Worksheet, depending on answers to the following validity questions.	anon n	un a j	51115	( )	011
VALIDITY QUESTIONS					
1. Was the <u>research question</u> clearly stated?		Y	N	UN	N
This is usually stated at end of the introduction and just before methods section.		S		C L	A
		Х		E A R	
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	X			
<i>This is often called the treatment and explained in the methods section.</i>					
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Х			
These are sometimes called the endpoints; the results section reports the outcomes, but					
this information should be in the methods section, too					
1.3 Were the target population and setting specified?	1.3	X			
The target population is group for whom findings may be applicable; look for this in the					
introduction and in the methods section			N		N
2. Was the <u>selection</u> of study subjects/patients free from bias?		E	0	N	A
		s		C L	
		v		E	
	-	Λ		R	
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1	X			
diagnostic or prognosis criteria), and with sufficient detail and without omitting					
The without do the study?					
The authors should give several points about the inclusion/exclusion criteria such as the					
age range of the subjects, disease condition (like hyperlipidemia) required for inclusion Exclusion criteria should be listed, too, although some are					
undarstood. East argumpla if the ages for inclusion are 18 to 70, the authors will					
understood. For example if the ages for inclusion are 10 to 70, the duitors will probably not specifically note that children and people over age 70 were					
excluded Most of the time however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD)					
2.2 Were criteria applied equally to all study groups?	2.2		-		X
2.2 Were health demographics and other characteristics of subjects described?	2.2	x	-		
There is usually a Table 1 summarizing demographics and characteristics at baseline	2.5				
Groups are not different if the P-Value is $> 0.05$ If there has been a previous					
paper describing the study population, that paper may be referenced and you					
would need to go back to the original publication to see that Table 1.					
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4	1	+	X	
The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may					
only say that the patients came from the same clinic from people who met the inclusion					
criteria.					1

3. W	Vere <u>study groups comparable</u> ?		Y E	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		ŝ	Ŭ	C	
	<i>Groups are</i> <u>not</u> different if the <i>P</i> -Value is $> 0.05$ .				E	
					A R	Х
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
	demographics) similar across study groups at baseline? See Table I for this - there					
	should be no significant differences across study groups in an intervention study.					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Х			
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
	prior study as a control group; that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control.					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4				Х
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	Х			
	reference standard (e.g. "gold standard")?					
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4.	Was method of handling withdrawals described?		Y	N O	UN	N A
			s	Ŭ	Ĉ	
					E	
					A R	Х
	4.1 Were follow up methods described and the same for all groups?	4.1				Х
	4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up.	4.2				Х
	attrition rate) and/or response rate (cross-sectional studies) described for each group?	-				
	(Follow up goal for a strong study is 80 %.)					
	This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
	note that on the worksheet (as a note in the results section or in the reviewer					
	comments at the very bottom)					
	4.3 Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Х			
	This information is often presented in a figure with # recruited. # enrolled (this is initial					
1	N), # remaining at end of study period (final N). Sometimes the reasons that subjects		1			
	withdrew or were dropped is given in the figure or in the text (results section).		1			
	4.4 Were reasons for withdrawals similar across groups?	4.4	1			Х
	If there is a large attrition from one group and not others. vou would want to look for a					-
	reason why; the answer to this question would then be no.					
· · · · · · · · · · · · · · · · · · ·						

	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	Х			
	test under study?					
	The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30\%$ .					
5.	Was <u>blinding</u> used to prevent introduction of bias?		Y E	N O	U N	N A
			s	-	C	
					E	
					A R	
					Х	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1				Х
	to treatment group, <u>as appropriate</u> ?					
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
	studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					
	an RCT, but obviously, the subjects and practitioners knew who was getting MNT					
	and who was not. Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2			Х	
	using an objective test, such as a lab value, this criterion is assumed to be met.)					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	<i>collecting the data).</i>					37
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Χ
	Establish who is a case and who is a control at the beginning of the study					
	<i>Establish who is a case and who is a control at the beginning of the study.</i>	5.5			v	
6	S.5 in diagnostic study, were test results blinded to patient instory and other test results?	3.3	Y	N		N
0.	comparison(s) described in detail? Wore intervening factors described?		E	0	N	А
	comparison(s) described in detail: were intervening factors described:		3		L	
			x		E A	
	(1 In DCT on other intercention trial more motorals described for all regiments at diado	6.1	Δ		R	v
	6.1 In KC1 of other intervention that, were protocols described for an regimens studied?	0.1	v			Λ
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2	Х			
	described?	6.2	v			
	nroduce a meaningful effect?	0.5	Λ			
	Use clinical judgment (a.g. 12 weeks is long arough for a dietary intervention to make a					
	difference in lab values for cholesterol: however 12 days would not be long					
	enough)					
	6.4 Was the amount of exposure and if relevant subject/national compliance measured?	64	x			
	How long did the treatment last? Did the patient follow directions?	0.1				
	6.5 Were co-interventions (e.g. ancillary treatments other therapies) described?	6.5				X
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)	0.0				
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					
	discussion section. It is likely there were no unplanned treatments, so a "no"					
	answer is not a problem overall.					
	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7				Х
	For a study to be valid and unbiased, it is important that this be yes.					
	6.8 In diagnostic study, were details of test administration and replication sufficient?	6.8	Х			
	Usually answer n/a for diet study.					

7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E	N A
			Х		A R	
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? Precision is reproducibility or repeatability.	7.5	X			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6		X		
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately?	8.1	X		R	
	<ul> <li>8.2 Were correct statistical tests used and assumptions of test not violated?</li> <li>You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).</li> </ul>	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)	8.3	Х			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4				X
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5				Х
	<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical</li> </ul>	8.6	X			

	significance is reported. Reducing cholesterol from 300 to 250 might be statistically significant, but clinically it is still abnormal.					
	<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li>Type II (β error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				X
9.	Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S	N O	U N C L E A	N A
	9.1 Is there a discussion of findings? Answer yes or no.	9.1	X		ĸ	
	9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	Х			
10.	<b>Is bias due to study's <u>funding or sponsorship</u> unlikely?</b> Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
	<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X			
	10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.	10.2	X			
SY.	MBOL					
MI If n min	NUS/NEGATIVE (-) aost (six or more) of the answers to the above validity questions are "no," the report should a us (-) symbol on the Evidence Quality Worksheet. UTRAL (0)	be design	ated	with	n a	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

right. Academy of Nutrition and Dietetics

Academy of Nutrition and Dietetics Evidence Analysis Library® Worksheet Template and Quality Criteria Checklist: Primary Research

Citation	Ferguson, T., Rowlands, A., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. International Journal of Behavioral Nutrition and Physical Activity, 12, 42. doi:10.1186/s12966-015-0201-9
Study Design	Validity study
Class	С
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\bigcirc$ (Neutral)
Research Purpose	To assess the concurrent validity of a selection of consumer-level accelerometer-based activity monitors compared to two research-grade accelerometers in free living conditions
Inclusion Criteria	Healthy adult participants over 18 years of age, living in metropolitan Adelaide, South Australia, and could ambulate without walking aids
Exclusion Criteria	Self-reported injury or illness affecting mobility
	Recruitment: Unclear, convenience sample
	Design: 21 healthy adult participants wore seven consumer-level activity
	monitors (Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike
	Fuelband, Striiv Smart Pedometer, and Withings Pulse) and two research-
	grade accelerometers/multi-sensor devices (BodyMedia SenseWear, and
	ActiGraph GT3X+) simultaneously for 48-hours. Participants went about
	their daily life in free-living conditions during data collection. Four
	physical activity parameters were measured, including step count,
Description of	moderate to vigorous physical activity (MVPA), sleep, and total daily
Study Protocol	energy expenditure (TDEE).
	Blinding used (if applicable): NA
	Intervention (if applicable): Seven consumer-level activity monitors
	(Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike Fuelband, Striiv
	Smart Pedometer, and Withings Pulse) were examined and chosen based
	on availability to the authors for purchase between February and August
	2013. The consumer-level devices were compared with two research
	grade tri-axial accelerometers/multi-sensor devices (BodyMedia
	SenseWear Model MF and ActiGraph GT3X+), which have accepted

reliability and validity as free-living measures of physical activity and sleep time. Participants attended an appointment at which demographic data (date of birth, sex, and dominant side) were obtained, with height and mass measured following standardized procedures. All nine devices were fitted to the participant in the following locations: SenseWear on the left upper arm; Fuelband, UP and Shine on the left wrist; GT3X+, One, Zip, Pulse and Striiv on the right side of the waist on an elasticized belt. Where consumer-level devices were designed for multiple wear locations, devices suitable for wrist wear were worn on the wrist; otherwise the device was worn on the waist. Placement order of the devices at the wrist and waist was randomized. Participants were instructed to leave all devices on simultaneously for approximately 48 hours (including sleep, but excluding showering) in order to capture a full overnight sleep episode as well as two 24-hours of activity data from midnight to midnight. The wear period was not limited to a particular period of the week (i.e. not restricted to weekdays only or weekends only), and no guidelines or restrictions on activity levels or sleep were provided, in order to ensure the study broadly represented free-living conditions. Participants were instructed how to turn sleep mode on and off for the relevant devices (Shine, Pulse, One, UP). Participants were not given access to any of the device software or account information and were also instructed not to turn off, modify, or change any device wear locations once fitted. Devices were collected after the 48-hour wear period for data collection. Data were extracted using the proprietary software for all consumer devices, in the same fashion that a consumer would utilize the software

Statistical Analysis: Participants' demographic data were analyzed descriptively. The validity of the consumer-level activity monitors relative to the research-grade devices for step count, MVPA, sleep, and TDEE was quantified using Bland-Altman analysis, median absolute difference, and Pearson's correlation. A priori power analyses were

	undertaken based on existing data on correlations among various research
	devices, which suggested that the correlation between consumer-level and
	research devices would be about 0.85. If the actual population correlation
	between consumer-level and research-grade devices was 0.85, then a
	target sample size of 21 would yield, in 95% of cases, a sample
	correlation between 0.65 and 0.94.
	Timing of Measurements: Data were collected continuously throughout
	the 48-hour wear period from all nine devices. However, data relating to
	physical activity were limited to the full calendar day (24-hour period
	midnight to midnight) following initialization, and data relating to sleep
	were limited to the first night of sleep (24-hour period midday to midday,
Data Collection	excluding naps) following initialization.
Summary	Dependent Variables: Physical activity parameters (step count, MVPA,
	sleep, and TDEE) according to Fitbit One, Fitbit Zip, Jawbone UP, Misfit
	Shine, Nike Fuelband, Striiv Smart Pedometer, and Withings Pulse
	Independent Variables: Free living conditions (e.g. participants' daily
	obligations, lifestyles, level of physical fitness, stress levels)
	Control Variables: Physical activity parameters (step count, MVPA, sleep,
	and TDEE) according to BodyMedia SenseWear and ActiGraph GT3X+
	Initial: 21 (10 Males 11 Females)
	Attrition (final N): 21
	Age: 20 to 59 years
Description of	Ethnicity: Unclear
Actual Data Sample	Other relevant demographics: All participants were right hand dominant
r i	Anthropometrics: Male BMI: $27.3 \pm 3.2 \text{ kg/m2}$ , female BMI: $25.5 \pm 5.2$
	kg/m2
	Location: Metropolitan Adelaide, South Australia
	Key Findings: All consumer-level activity monitors measured steps, and
Summary of Results	correlations with reference devices were very strong ( $r = 0.94-0.99$ ).
ixesuits	Bland-Altman analyses suggested that three of the activity monitors

	slightly over-counted (Striiv, Zip, One) while four under-counted
	(Fuelband, Shine, Up, Pulse). Five of the activity monitors (Striiv, Shine,
	Up, Zip, One) were considered to measure a parameter similar or
	equivalent to MVPA time. Correlations between readings from the
	activity monitors and reference devices ranged from weak to strong (r =
	0.52-0.91). Bland-Altman analyses showed large differences between the
	mean values reported. For example, the Shine under-counted (mean =
	53.3 min of MVPA compared to reference device $(GT3X+)$ mean = 58.5
	min), while the Striiv over-counted (mean = 249 min of MVPA compared
	to reference device (GT3X+)). Of the five activity monitors (Shine, Up,
	Pulse, Zip, One) that measured TDEE, correlations with the reference
	devices were moderate to strong ( $r = 0.74$ - 0.81). Bland-Altman analyses
	suggest all activity monitors considerably underestimated TDEE
	compared to the reference device (SenseWear, mean = 3005 kcal),
	ranging from 475 kcal (One) to 898 kcal (UP). Of the four activity
	monitors (Shine, Up, Pulse, One) that measured minutes of sleep, all
	correlated strongly with the reference device ( $r = 0.82-0.92$ ). Bland-
	Altman analyses showed all activity monitors overestimated minutes of
	sleep, most notably, the Shine (mean = 44 min) compared to reference
	device (SenseWear) mean = 424 min).
	Other Findings:
	In free-living conditions, the consumer-level activity monitors showed
	strong validity for the measurement of steps and sleep duration, and
	moderate-to-strong valid for measurement of TDEE and MVPA. Median
Author	absolute differences were generally modest for sleep and steps, moderate
Conclusion	for TDEE, and large for MVPA. Validity for each construct ranged
	widely between activity monitors, with the Fitbit One, Fitbit Zip, and
	Withings Pulse being the strongest performers.
Reviewer	Strengths: zero percent attrition, the use of numerous consumer and
Comments	reference devices, testing the devices in free-living conditions as they are

	designed for, and examining several different physical activity variables
	collected by the devices
	Limitations: participant recruitment was not discussed, blinding was not
	utilized, and validity may vary if activity monitors are worn in locations
	other than the hip or wrist
Funding Source	Unclear, authors declared no competing interests

### Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
4	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
U U	exceptionally week

Select a rating from the drop-down menu

		uroj	p-uown menu ♥	
R	Relevance Questions			
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear	
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes	
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No	
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes	

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Validity Questions			
1.	Was the <u>research question</u> clearly stated?	1	Yes
	1.1. Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	Yes
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3. Were the target population and setting specified?	1.3	Yes
2. Was the <u>selection</u> of study subjects/patients free from bias?		2	Yes
	progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes
	2.2. Were criteria applied equally to all study groups?	2.2	N/A
	2.3. Were health, demographics, and other characteristics of subjects described?	2.3	Yes
	population?	2.4	Unclear

3.	Were	study groups comparable?	2	NI/A
	3.1.	Was the method of assigning subjects/patients to groups described and unbiased? (Method of randomization identified if RCT)	3	IN/A
	3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	N/A
	3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	3.2	N/A
	3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.3	Yes
	3.5.	appropriate adjustments in statistical analysis? If case control study, were potential confounding factors comparable for cases	3.4	N/A
		and controls? (If case series or trial with subjects serving as own control, this criterion is not applicable. Criterion may not be applicable in some cross-	3.5	N/A
	3.6.	If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	Yes
4.	Was n	nethod of handling withdrawals described?	4	N/A
	4.1.	Were follow up methods described and the same for all groups?		
	4.2.	Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow	4.1	N/A
		each group? (Follow up goal for a strong study is 80%.)	4.2	N/A
	4.3.	Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Yes
	4.4.	Were reasons for withdrawals similar across groups	4 4	N/A
	4.5.	If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4.5	Yes
-	Wac b	linding used to provent introduction of bias?		
5.	vvas <u>D</u>	intuing used to prevent introduction of blas:	5	Linglage
5.	5.1.	In intervention study, were subjects, clinicians/practitioners, and investigators	5	Unclear
5.	5.1.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is	5 5.1	Unclear N/A
5.	5.1. 5.2.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)	5 5.1 5.2	Unclear N/A Unclear
5.	5.1. 5.2. 5.3.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	5 5.1 5.2 5.3	Unclear N/A Unclear N/A
5.	5.1. 5.2. 5.3. 5.4.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5       5.1       5.2       5.3       5.4	Unclear N/A Unclear N/A N/A
5.	5.1. 5.2. 5.3. 5.4. 5.5.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results?	5 5.1 5.2 5.3 5.4 5.5	Unclear N/A Unclear N/A N/A Unclear
5.	• was <u>b</u> 5.1. 5.2. 5.3. 5.4. 5.5. <b>Were</b>	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results?	5         5.1         5.2         5.3         5.4         5.5         6	Unclear N/A Unclear N/A N/A Unclear Yes
5.	5.1. 5.2. 5.3. 5.4. 5.5. Were compa	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? <b>intervention/therapeutic regimens/exposure factor or procedure and any</b> <b>arison(s) described in detail? Were <u>intervening factors</u> described?</b>	5 5.1 5.2 5.3 5.4 5.5 6 6	Unclear N/A Unclear N/A N/A Unclear Yes N/A
6.	• Was <u>b</u> 5.1. 5.2. 5.3. 5.4. 5.5. • Were compa 6.1.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? Intervention/therapeutic regimens/exposure factor or procedure and any arison(s) described in detail? Were intervening factors described? In RCT or other intervention trial, were protocols described for all regimens studied?	5         5.1         5.2         5.3         5.4         5.5         6         6.1	Unclear N/A Unclear N/A N/A Unclear Yes N/A
6.	•••as <u>b</u> 5.1. 5.2. 5.3. 5.4. 5.5. ••••••••••••••••••••••••••••	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? <b>intervention/therapeutic regimens/exposure factor or procedure and any</b> <b>arison(s) described in detail? Were intervening factors described?</b> In RCT or other intervention trial, were protocols described for all regimens studied? In observational study, were interventions, study settings, and	5         5.1         5.2         5.3         5.4         5.5         6         6.1         6.2	Unclear N/A Unclear N/A N/A Unclear Yes N/A Yes
6.	Was <u>b</u> 5.1.           5.2.           5.3.           5.4.           5.5.           Were compa           6.1.           6.2.           6.3.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? <b>intervention/therapeutic regimens/exposure factor or procedure and any</b> <b>arison(s) described in detail? Were intervening factors described?</b> In observational study, were interventions, study settings, and clinicians/provider described? Was the intensity and duration of the intervention or exposure factor sufficient	5         5.1         5.2         5.3         5.4         5.5         6         6.1         6.2         6.3	Unclear N/A Unclear N/A N/A Unclear Yes N/A Yes Yes
6.	Was <u>b</u> 5.1. 5.2. 5.3. 5.4. 5.5. Were compa 6.1. 6.2. 6.3.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? intervention/therapeutic regimens/exposure factor or procedure and any arison(s) described in detail? Were intervening factors described? In RCT or other intervention trial, were protocols described for all regimens studied? In observational study, were interventions, study settings, and clinicians/provider described? Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?	5         5.1         5.2         5.3         5.4         5.5         6         6.1         6.2         6.3         6.4	Unclear N/A Unclear N/A N/A Unclear Yes N/A Yes Yes Yes Yes
6.	Was <u>b</u> 5.1.         5.2.         5.3.         5.4.         5.5.         Were         compa         6.1.         6.2.         6.3.         6.4.	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? intervention/therapeutic regimens/exposure factor or procedure and any arison(s) described in detail? Were intervening factors described? In RCT or other intervention trial, were protocols described for all regimens studied? In observational study, were interventions, study settings, and clinicians/provider described? Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect? Was the amount of exposure and, if relevant, subject/patient compliance measured?	5         5.1         5.2         5.3         5.4         5.5         6         6.1         6.2         6.3         6.4         6.5	Unclear N/A Unclear N/A N/A Unclear Yes N/A Yes Yes Yes Yes Yes
6.	Was <u>b</u> 5.1. 5.2. 5.3. 5.4. 5.5. Were compa 6.1. 6.2. 6.3. 6.4. 6.5. 6.6	In intervention study, were subjects, clinicians/practitioners, and investigators blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.) In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded? In case control study, was case definition explicit and case ascertainment not influenced by exposure status? In diagnostic study, were test results blinded to patient history and other test results? <b>intervention/therapeutic regimens/exposure factor or procedure and any</b> <b>prison(s) described in detail? Were intervening factors</b> described? In RCT or other intervention trial, were protocols described for all regimens studied? In observational study, were interventions, study settings, and clinicians/provider described? Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect? Was the amount of exposure and, if relevant, subject/patient compliance measured? Were co-interventions (e.g., ancillary treatments, other therapies) described? Were extra or upplaned treatments described?	5         5.1         5.2         5.3         5.4         5.5         6         6.1         6.2         6.3         6.4         6.5         6.6	Unclear N/A Unclear N/A N/A Unclear Yes N/A Yes Yes Yes Yes N/A N/A

	6.8. In diagnostic study, were details of test administration and replication sufficient?	6.8	Yes
7.	Were outcomes clearly defined and the measurements valid and reliable?	7	Yes
	7.1. Were primary and secondary endpoints described and relevant to the question?	7.1	Yes
	7.2. Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A
	7.3. Was the period of follow-up long enough for important outcome(s) to occur?	7.3	N/A
	7.4. Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?	7.4	Yes
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	No
	7.7. Were the measurements conducted consistently across groups?	7.0	
_		1.1	N/A
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?	8	Yes
	8.1. Were statistical analyses adequately described the results reported	8.1	Yes
	appropriately?	0 7	Var
	8.2. Were correct statistical tests used and assumptions of test not violated?	0.2	res
	<ul> <li>8.3. Were statistics reported with levels of significance and/or confidence intervals?</li> <li>8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there</li> </ul>	8.3	Yes
	an analysis of outcomes for those maximally exposed or a dose-response analysis)?	8.4	N/A
	8.5. Were adequate adjustments made for effects of confounding factors that	8.5	N/A
	might have affected the outcomes (e.g., multivariate analyses)? 8.6. Was clinical significance as well as statistical significance reported?	8.6	Yes
	8.7. If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A
9.	Are <u>conclusions supported by results</u> with biases and limitations taken into	9	Yes
	consideration?	9.1	Yes
	9.1. Is there a discussion of findings?	9.2	Yes
10.	Is bias due to study's funding or sponsorship unlikely?	10	Ves
	10.1. Were sources of funding and investigators' affiliations described?	10 1	Ves
	10.2. Was there no apparent conflict of interest?	10.1	Yes
М			1

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

#### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

## **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Gomersall, S., Ng, N., Burton, N., Pavey, T., Gilson, N., & Brown, W. (2016). Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. <i>Journal of Medical Internet Research</i> , <i>18</i> (9), e239.
	doi:10.2196/jmir.5531
Study design: Use algorithm – RCT, cohort, etc	Validity study
Study Class (A,B,C,D)	С
Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø) This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).	NEUTRAL (ø)
	Purpose/Population Studied/Practice Studied
<b>Research purpose:</b> What is the research question being investigated in the study?)	To compare Fitbit One and Jawbone UP estimates of steps, moderate-to-vigorous physical activity (MVPA), and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context
Inclusion criteria: requirements for study eligibility	Healthy, ambulatory, adult participants between 18 and 65 years of age, have accumulated less than 150 minutes of MVPA in the past week (assessed using the Active Australia Survey), and own or have access to a mobile phone compatible with both the Fitbit One and Jawbone UP
<b>Exclusion criteria</b> (conditions that make individual ineligible)	Unclear
Recruitment	Convenience sampling at three campuses of a large Australian metropolitan university via an email advertisement to staff that included study information and participant eligibility criteria
<b>Blinding used:</b> some of the persons involved are prevented from knowing certain information that might lead to conscious or unconscious bias on their part, invalidating the results	Data were extracted via the users' accounts and entered into an Excel spreadsheet by a research assistant
<b>Description of study protocol</b> <i>What happened in the study?</i>	On two occasions for seven days each, participants wore an ActiGraph GT3X+ accelerometer on their right hip and either a hip-worn Fitbit One or wrist-worn Jawbone UP activity tracker. Daily estimates of steps and very active minutes were derived from the Fitbit One, and steps, active time, and longest idle time were derived from the Jawbone UP. Daily estimates of steps, MVPA, and longest sedentary bout were derived from the corresponding days of ActiGraph data.
Intervention: Describe interventions, regimens, risk factors, or procedures studied.	Data were collected as part of a larger, 12-week physical activity intervention study that included three groups that were randomly allocated to wear a Fitbit One, Jawbone UP, or standard pedometer. Demographic and anthropometric data were collected at baseline. Data for this substudy were collected at mid- and post- intervention when participants concurrently wore an ActiGraph GT3X+ accelerometer. On these two occasions for seven days each, participants wore an ActiGraph GT3X+ accelerometer on their right hip and either a hip-worn Fitbit One (n=14) or wrist-worn Jawbone UP (n=15) activity tracker. Participants were instructed to wear the devices during waking hours, removing them for water-based activities or contact sports, but were not required to keep wear logs in order to
	improve the free-living fidelity of the devices over the 12-week intervention.
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
	Participants were able to input activity sessions, such as swimming or contact
	sports, through the "log workout" function in the Jawbone UP app and the "track
	exercise" feature in the Fitbit app. Daily estimates of steps and very active minutes
	were derived from the Fitbit One, and steps, active time, and longest idle time were
	derived from the Jawbone UP Daily estimates of steps MVPA and longest
	sedentary hout were derived from the corresponding days of ActiGraph data Data
	were extracted via the users' accounts and entered into an Excel spreadsheet by a
	research assistant.
Statistical analysis: List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Descriptive statistics (n, mean, standard deviation, and prevalence) were calculated for demographic and physical measures. Absolute agreement was examined using intraclass correlation coefficients (ICC) and 95% confidence intervals. Correlation was assessed using Pearson correlation coefficient or Spearman rank correlation coefficient when data were non-normally distributed with 95% confidence intervals. Bland-Altman plots were used to examine the differences between all outcomes, with mean bias and 95% limits of agreement reported. Linear regression was used to examine whether mean difference and limits of agreement varied across mean values of Fitbit One or Jawbone UP and ActiGraph outcomes. Cohen's kappa statistic was used to assess the agreement between devices for classification of active versus inactive based first on achieving 10,000 steps or more per day (default step goal on both devices) and second on achieving 30 minutes per day or more of MVPA (comparable with public health guidelines).
	<i>P</i> values were based on two-sided tests and were considered statistically significant
	at $P < 05$ . Post hoc power calculations determined that a sample size of N=289
	daily comparisons would detect correlations as low as 17 with 80% power and 5%
	alpha
Timing of measurements: <i>when</i>	Data was collected continuously as participants were instructed to wear either a
outcomes were measured; usually	Fitbit One or Jawbone UP activity tracker every day during the 12-week physical
baseline and one or more later	activity intervention. Daily estimates of steps, active time, and longest idle time
times	were extracted from the users' accounts. Participants concurrently wore an
	ActiGraph GT3X+ accelerometer at mid- and post-intervention for seven days.
	Daily estimates of steps, MVPA, and longest sedentary bout were derived from the
	corresponding days of ActiGraph data.
Dependent variables: <i>outcomes</i>	Steps and very active minutes according to the Fitbit One
that are measured or registered;	Steps, active time, and longest idle time according to the Jawbone UP
variable whose change or	
different states the researcher	
wants to understand, explain, or	
nredict	
Independent variables	Free living context (e.g. participants' daily obligations, lifestyles, level of physical
(intervention or procedure: this	fitness, stress levels)
variable can be manipulated: a	
variable whose effect upon the	
dependent variable one is trying	
to understand)	
Control Variables	Steps MVPA and longest sedentary bout according to the ActiGraph GT3X+
Examples: 1) multivariate logistic	accelerometer
regression controlled for age	
RML albumin:	
2) usual care: 3) isocaloric diet	
etc	
<b>Initial n</b> (e.g. 731 (298 males	32 participants (only 29 provided valid data for the current analyses)
433 females))	22 paraopants (only 2) provided value data for the eartent anaryses)
Record number that entered	
study – not the number screened.	

Final n (attrition)	29 participants, 3 males and 26 females
number of subjects that completed	
Age usually mean or range	Mean age: 39.6, SD: 11.0 years
Ethnicity (if given)	Unclear
Other relevant demographics:	25 participants (86%) completed tertiary education
demographics describe the	
population (students, athletes, etc)	
Anthropometrics: e.g. were	Mean BMI: 25.9, SD: 5.0 kg/m2
groups same or afferent on important physical measures	
(BMI, fitness level)	
<b>Location:</b> <i>Where did the study</i>	Australia
take place? City or country	
results including quantitative data and statistics. Include statistical significance: P-values, confidence intervals (CI), relative risk (RR), odds ratios (OR), likelihood ratio, number needed to treat power	Fitbit One ( $r=.85$ for steps and $\rho=.80$ for MVPA) than for Jawbone UP ( $r=.75$ for steps and $\rho=.75$ for MVPA). The correlation between Jawbone UP longest idle time and ActiGraph longest sedentary bout was poor ( $\rho=.19$ ). Absolute agreement (ICC) was acceptable for ActiGraph and Fitbit One steps (.90) and MVPA (.72) and Jawbone UP steps (.79). However, agreement was weak between ActiGraph and Iawbone UP estimates of MVPA (.56) and longest idle time (.08). For the
number needed to treat, power analysis if available.	and Jawbone UP estimates of MVPA (.56) and longest idle time (.08). For the estimation of steps, 95% limits of agreement were unbiased for both devices, although limits were wider for Jawbone UP than for Fitbit One (5290 and 3567 steps/day). When absolute values were calculated, both devices overestimated steps (Fitbit One: mean bias 767, 95% limits of agreement –2800 to 4334; Jawbone UP: mean bias 1178, 95% limits of agreement –4112 to 6468). For the estimation of MVPA, bias was evident for both the mean difference and the limits of agreement for both the Fitbit One and the Jawbone UP. When absolute values were calculated, the Fitbit One underestimated MVPA by a mean 19.2 minutes/day (95% limits of agreement 5.8-65). For the estimation of longest sedentary bout, the limits of agreement steps of 38.1 minutes/day (95% limits of agreement were unbiased but wide (mean difference ±88 minutes), varying by up to 150% of the mean estimate according to ActiGraph. Using the criterion of at least 10,000 steps per day, agreement between the Fitbit One and ActiGraph for the classification of active versus inactive was substantial ( $\kappa$ =.68, <i>P</i> <.001). The Fitbit One correctly classified 95% of days as active and 79% of days as inactive. Agreement between the Jawbone UP and ActiGraph was fiair ( $\kappa$ =.39, <i>P</i> <.001). The Fitbit One correctly classified 100% of days as active and 100% of days as inactive. Agreement between the Jawbone UP and ActiGraph was slight ( $\kappa$ =.14, <i>P</i> =.001). The Jawbone UP correctly classified 100% of days as active and 12% of days as inactive.
Author conclusion: <i>paraphrase</i>	The findings reported in this study suggest that both activity trackers have utility
that stated by study author in body of the report or abstract	for counting steps in free-living settings, with both devices overestimating daily steps by only 5% to 15% compared with ActiGraph (Fitbit One: 8%; Jawbone UP: 14%). Both devices were less accurate measuring MVPA than steps, with correlations of .56 to .80 for both devices against ActiGraph data. Despite reasonable correlations, the Fitbit One underestimated MVPA by 46%, while the Jawbone UP overestimated MVPA by 50%. Findings indicate that the validity of the Jawbone UP measure of sedentary behavior (longest idle time) compared with ActiGraph-determined "longest sedentary bout" was poor. Both devices accurately
	classified more than 80% of the sample days as active or inactive based on the

	10,000 steps criterion; however, days were frequently misclassified for meeting public health guidelines of 30 minutes/day of MVPA. Due to modest accuracy and systematic bias, both activity trackers are better suited as self-monitoring tools (e.g.
	for the public consumer or in behavior change interventions) rather than for
	evaluation of research outcomes. The outcomes that relate to sedentary behavior
	and MVPA should be used with caution for both consumers and researchers alike.
<b>Reviewer comments:</b> Note strengths and limitations of study; identify concerns that affect study validity and generalizability— your comments should be italicized)	Strengths: concurrent assessment of two popular brands of activity trackers on the market and two popular wear locations (wrist and waist), large number of daily observations for comparison, free-living setting which improves ecological validity and takes previous laboratory studies into a real-world setting, sample had good wear compliance, and the thorough evaluation of systematic bias Limitations: predominantly female, healthy, middle-aged sample which limits the generalizability of the findings, and the study could not control for wear time of the activity trackers which may explain some of the large absolute differences between the devices and the ActiGraph Funding source: Start-Up Grant from The University of Queensland. Drs. Gomersall and Pavey were supported by an Australian National Health and Medical Research Council (NHMRC) program grant (NHMRC no: 569940).
	Medical Research Council (NHMRC) program grant (NHMRC no: 569940). Authors declared no conflicts of interest.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet) Gomersall, S., Ng, N., Burton, N., Pavey, T., Gilson, N., & Brown, W. (2016). Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. Journal of Medical Internet Research. 18(9), e239. doi:10.2196/imir.5531		Y E S	N O	U N C L E A R	N A
<ol> <li>Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some epidemiological studies)</li> </ol>	1			Х	
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/ population group would care about?	2	Х			
3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3		Х		
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "yes", the report is eligible for design the Evidence Quality Worksheet, depending on answers to the following validity questions.	nation w	ith a	plus	: (+)	on
VALIDITY QUESTIONS	-	V			N
<b>1. Was the <u>research question</u> clearly stated?</b> <i>This is usually stated at end of the introduction and just before methods section.</i>		Y E S	C	D N C L E	A
		Х	C .	A R	
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified? <i>This is often called the treatment and explained in the methods section.</i>	1.1	Х	[		
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated? These are sometimes called the endpoints; the results section reports the outcomes, but this information should be in the methods section, too	1.2	Х			
1.3 Were the target population and setting specified? <i>The target population is group for whom findings may be applicable; look for this in the</i> <i>introduction and in the methods section</i>	1.3	Х	5		
2. Was the <u>selection</u> of study subjects/patients free from bias?		Y E S	N C	U N C L A R X	N A
<ul> <li>2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?</li> <li>The authors should give several points about the inclusion/exclusion criteria such as the age range of the subjects, disease condition (like hyperlipidemia) required for inclusion. Exclusion criteria should be listed, too, although some are understood. For example if the ages for inclusion are 18 to 70, the authors will probably not specifically note that children and people over age 70 were excluded. Most of the time, however, subjects may be excluded for certain characteristics such as being pregnant or having some disease (like CHD).</li> </ul>	2.1				
2.2 Were criteria applied equally to all study groups?	2.2	X			$\bot$
<ul> <li>2.3 Were health, demographics, and other characteristics of subjects described? There is usually a Table 1 summarizing demographics and characteristics at baseline. Groups are <u>not</u> different if the P-Value is &gt; 0.05. If there has been a previous paper describing the study population, that paper may be referenced and you would need to go back to the original publication to see that Table 1.</li> <li>2.4 Were the subjects/patients in a representative sample of the relevant population?</li> </ul>	2.3				T
2.4 were the subjects/patients in a representative sample of the relevant population?	2.4				•

		_		-	-
The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion criteria					
3 Were study groups comparable?		Y	Ν	U	N
There is usually a Table 1 summarizing demographics and characteristics at baseline		ES	0	N C	Α
Groups are not different if the $P_{-}$ Value is > 0.05		5		Ľ	
$\frac{1}{1000} \frac{1}{1000} \frac{1}{1000$				E A	
				R V	
2.1 Was the method of assigning subjects/nationts to groups described and unbiased?	2 1				
(Method of randomization identified if PCT)	5.1			Λ	
In a study, the authors may tall how the subjects were assigned to a group (a g					
In a strong study, the authors may tell now the subjects were assigned to a group (e.g.					
Look for instances that show higs: for example Longe read a study where patients					
Look for instances that show olds, for example 1 once read a study where patients					
disliked their supplement, they were allowed to change groups this is not unbiased	e				
2.2 Ware distribution of disease status, prognostic feature, and other factors (a.g.	2.2			v	
demographics) similar across study groups at baseline? See Table I for this there	3.2			Λ	
should be no significant differences games study groups in an intervention study					
2.2 Ware concurrent controls used? (Concurrent proferred over historical controls)	2.2	v			
S.5 were concurrent controls used? (Concurrent preferred over instorical controls.)	5.5	Λ			
Most RC1s use a concurrent control group. Occasionally an intervention study will use a					
prior study as a control group, that is an example of a historical control. That is not					
as strong a research design as use of concurrent control group. A crossover study					
2.4. If exhapt study or cross sectional study, were groups comparable on important	2.4				v
3.4 If conort study of cross-sectional study, were groups comparable on important	5.4				Λ
appropriete adjustments in statistical analysis?					
The groups in a cohort or cross sectional study should not be different from each other:					
if they are a strong study will utilize statistical techniques such as multivariate					
if they are, a shong study will utilize statistical techniques such as multivariate analyses to remove the variance due to the group differences. Look for this					
information in the statistics and results sections					
2.5. If case control study, were notential confounding factors comparable for cases and	3.5				v
controls? If case series or trial with subjects serving as own control, this criterion is	5.5				Λ
not applicable. Criterion may not be applicable in some cross-sectional studies					
Subjects are generally matched for age gender ate. Look for this in the statistical					
description and results sections					
3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	x			
reference standard (e.g. "gold standard")?	5.0	11			
Example: comparing body fat analysis method with underwater weighing (gold					
standard) In studies trying to determine the best equation (like Mifflin-St. Jeor or					
Harris-Renedict) to predict energy needs a gold standard measure of RFF (Indirect					
Calorimetry) is used					
4 Was method of handling withdrawals described?		Y	N	U	N
i. Was method of handning <u>withdrawais</u> described.		E S	0	N C	A
				L	
				A	X
4.1 Were follow up methods described and the same for all groups?	4.1	-		R	v
4.1 were follow up methods described and the same for an groups?	4.1				
4.2 was the number, characteristics of withdrawais (i.e. dropouts, lost to follow up,	4.2				Х
attrition rate) and/or response rate (cross-sectional studies) described for each group?					
(rollow up goal for a strong study is $80\%$ .)					
I have shown a be journed in the results section. If there is attrition $> 20\%$ , it is important to					
note that on the worksheet (as a note in the results section or in the reviewer					
4.2 Wore all approximate (in the animial counted for 9	4.2	-	v		
4.5 were an enroned subjects/patients (in the original sample) accounted for?	4.3		Λ		

	This information is often presented in a figure with # recruited, # enrolled (this is initial					
	N), # remaining at end of study period (final N). Sometimes the reasons that subjects					
	<i>withdrew or were dropped is given in the figure or in the text (results section).</i>	4.4	<u> </u>		-	v
	4.4 Were reasons for withdrawars similar across groups? If there is a large attrition from one group and not others, you would want to look for a	4.4				Λ
	<i>if more is a large all monormone group and not onless, you would want to took for a reason why: the answer to this question would then be no</i>					
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	X			
	test under study?					
	The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					
	<i>BMI</i> was $>$ 35 but bioimpedance analyzer indicated body fat $<$ 30%.					
5.	Was <u>blinding</u> used to prevent introduction of bias?		Y E	N O	U N	N A
			s		C L	
					E A	
			Х		R	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1				Х
	to treatment group, as appropriate?					
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
	studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					
	an RCT, but obviously, the subjects and practitioners knew who was getting MNT					
	and who was not. Therefore, you would not answer question 5.1 NO. It was					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2	v			
	using an objective test such as a lab value this criterion is assumed to be met )	5.2	1			
	Answer ves, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
		5.4	_		_	v
	5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4		$\square$		X
	<ul><li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li><li>Establish who is a case and who is a control at the beginning of the study.</li></ul>	5.4				X
	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> </ul>	5.4	x			X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any</li> </ul>	5.4	X	N	U	X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><u>Establish who is a case and who is a control at the beginning of the study.</u></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> </ul>	5.4	X Y E S	N O	U N C	X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><u>Establish who is a case and who is a control at the beginning of the study.</u></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> </ul>	5.4	X Y E S	N O	U N C L E	X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><u>Establish who is a case and who is a control at the beginning of the study.</u></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> </ul>	5.4	X Y E S X	N O	U N C L E A R	X N A
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><u>Establish who is a case and who is a control at the beginning of the study.</u></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> </ul>	5.4 5.5 6.1	X Y E S X	N O	U N C L E A R	X N A
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider</li> </ul>	5.4 5.5 6.1 6.2	X Y E S X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> </ul>	5.4 5.5 6.1 6.2	X Y E S X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to</li> </ul>	5.4 5.5 6.1 6.2 6.3	X Y E S X X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> </ul>	5.4         5.5         6.1         6.2         6.3	X Y S X X X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a</li> </ul>	5.4         5.5         6.1         6.2         6.3	X Y S X X X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long</li> </ul>	5.4         5.5         6.1         6.2         6.3	X Y S X X X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> </ul>	5.4 5.5 6.1 6.2 6.3	X Y E S X X X	N O	U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li><i>Establish who is a case and who is a control at the beginning of the study.</i></li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?</li> </ul>	5.4         5.5         6.1         6.2         6.3         6.4	X Y S X X X X X		U N C L E A R	X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured? How long did the treatment last? Did the patient follow directions?</li> <li>6.5 Were co-interventions (e.g., ancillary treatments other therapice) described?</li> </ul>	5.4 5.5 6.1 6.2 6.3 6.4	X Y S X X X X X		U N C L E A R	X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured? How long did the treatment last? Did the patient follow directions?</li> <li>6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?</li> </ul>	5.4         5.5         6.1         6.2         6.3         6.4         6.5	X Y S X X X X X X X			X A X
6.	<ul> <li>5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?</li> <li>Establish who is a case and who is a control at the beginning of the study.</li> <li>5.5 In diagnostic study, were test results blinded to patient history and other test results?</li> <li>Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?</li> <li>6.1 In RCT or other intervention trial, were protocols described for all regimens studied?</li> <li>6.2 In observational study, were interventions, study settings, and clinicians/provider described?</li> <li>6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?</li> <li>Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a difference in lab values for cholesterol; however, 12 days would not be long enough)</li> <li>6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured? How long did the treatment last? Did the patient follow directions?</li> <li>6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?</li> <li>6.4 Were extra or unplanned treatments described?</li> </ul>	5.4 5.5 6.1 6.2 6.3 6.4 6.4 6.5	X Y E S X X X X X		U N C L E A R	X A X

	The text may not describe any unplanned treatments. If yes, it would likely be in the discussion section. It is likely them were no unplanned treatments as a "no"					
	answer is not a problem overall					
	6.7 Was the information for 6.4.6.5.6.6 and 6.7 assessed the same way for all groups?	67	x			
	For a study to be valid and unbiased, it is important that this be yes.	0.7	11			
	6.8 In diagnostic study, were details of test administration and replication sufficient?	6.8	Х			
	Usually answer n/a for diet study.					
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y	N	U N	N A
			ŝ	Ŭ	Ĉ	
					E	
			Х		A R	
	7.1 Were primary and secondary endpoints described and relevant to the question?	7.1	Х			
	Primary endpoint –main result measured at the end of a study to see if the treatment					
	worked. The primary endpoint is decided at the beginning of the study.					
	Secondary endpoint - not as important as the main results; not usually analyzed if the					
	primary endpoint is not statistically significant.					
	7.2 Were nutrition measures appropriate to question and outcomes of concern?	7.2				Х
	Clinical judgment required: weight loss, changes in energy intake are relevant to MNT;					
	Sometimes there are no nutrition measures and you should answer N/A.					
	7.3 Was the period of follow-up long enough for important outcome(s) to occur?	7.3				Х
	Clinical judgment required: was there enough time?	7.4	v			
	/.4 Were the observations and measurements based on standard, valid, and reliable data	7.4	А			
	Check that surveys were validated					
	7.5 Was the measurement of effect at an appropriate level of precision?	7.5	v			
	Precision is reproducibility or repeatability	7.5	Λ			ļ
	7.6 Were other factors accounted for (measured) that could affect outcomes?	7.6		X		
	Other factors are sometimes covered in the discussion of the strengths/limitations of the	/.0				
	study.					
	7.7 Were the measurements conducted consistently across groups?	7.7	Х			
8.	Was the statistical analysis appropriate for the study design and type of outcome		Y	N	U	N
	indicators?		E S	0	C N	А
					L E	
			Х		A R	
	8.1 Were statistical analyses adequately described and the results reported appropriately?	8.1	Х			
	There should be a discussion of the statistics in the methods section.					
	8.2 Were correct statistical tests used and assumptions of test not violated?	8.2	Х			
	You will get better at this the more papers you abstract.EAL abstractors are expected to					
	have some statistical and research training (minimum of master's degree).					
	8.3 Were statistics reported with levels of significance and/or confidence intervals?	8.3	Х			
	( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)					
	8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an	8.4				Х
	analysis of outcomes for those maximally exposed or a dose-response analysis)?					
	Intent to treat- analysis is based on the original treatment intent, not the treatment					
	ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the					
	weeks in the study). The analyses are done using all the subjects in the study, not just					
	the ones who completed it. This is done in order to avoid effects of dropout that can					
	be a threat to randomization. Intent-to-treat analysis of outcomes applies to any interpretion study. If intent to treat analysis and the it will be marticed in the					
	be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any</i> <i>intervention study</i> . If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it intent to treat					
	be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any</i> <i>intervention study</i> . If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done					
	<ul> <li>be a threat to randomization. Intent-to-treat analysis of outcomes applies to any intervention study. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> <li>8 5 Were adequate adjustments made for effects of confounding factors that might have</li> </ul>	8.5				X

Multivariate analyses are used to adjust or control for other variables (age, sex,					
smoking, etc). Assumes data is valid and reduces a larger number of variables to a					
smaller number. Just answer yes or no that multivariate analyses were used.					
8.6 Was clinical significance as well as statistical significance reported?	8.6	Х			
Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was					
reduced from 229.2 $\pm$ 158 to 181.3 $\pm$ 16.3 (P<0.001); This includes: statistical					
significance (P-value) and clinical significance (compare to standard of $< 200$					
mg/do for normal cholesterol). A problem can occur when only statistical					
significance is reported. Reducing cholesterol from 300 to 250 might be statistically					
significant, but clinically it is still abnormal.					
8.7 If negative findings, was a power calculation reported to address type 2 error?	8.7			X	
<i>Type II (<math>\beta</math> error is a false negative that happens when the investigators fail to reject the</i>					
<u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say					
something like "a sample size of n=xx is needed to provide 80% power."					
9. Are <u>conclusions</u> supported by results with biases and limitations taken into		Y E	N O	U N N A	
consideration?		s		C L	
				Ē	
		Х		A R	
9.1 Is there a discussion of findings?	9.1	Х			
Answer yes or no.					
9.2 Are biases and study limitations identified and discussed?	9.2	Х			
This will be in the discussion of finding section that follows the results					
10. Is bias due to study's <u>funding or sponsorship</u> unlikely?		Y E	N O	U N N A	
Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		s		C	
				E	
		Х		A R	
10.1 Were sources of funding and investigators' affiliations described?	10.1	Х			_
• Look just under the abstract, or					
• The funding may be acknowledged at the end of the paper					
• Just because the work was funded by industry does not mean the study was biased.					
10.2 Was there no apparent conflict of interest?	10.2	Х			
If an investigator is testing a piece of equipment, process or drug that s/he developed, it					
could be a conflict of interest.					
SYMBOL			. <u> </u>		
MINUS/NEGATIVE (-)					
If most (six or more) of the answers to the above validity questions are "no," the report should be	be design	ated	with	а	
minus (-) symbol on the Evidence Quality Worksheet.					
NEUTRAL (Ø)					
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is except.	ionally st	rong	the		
report should be designated with a neutral (ø) symbol on the Evidence Quality Worksheet.					
PLUS/POSITIVE (+)					

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Academy of Nutrition and Dietetics Evidence Analysis Library® Worksheet Template and Quality Criteria Checklist: Primary Research

Citation	Gomersall, S., Ng, N., Burton, N., Pavey, T., Gilson, N., & Brown, W. (2016). Estimating Physical Activity and Sedentary Behavior in a Free- Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. Journal of Medical Internet Research, 18(9), e239. doi:10.2196/jmir.5531
Study Design	Validity study
Class	C
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\boxtimes \otimes$ (Neutral)
Research Purpose	To compare Fitbit One and Jawbone UP estimates of steps, moderate-to- vigorous physical activity (MVPA), and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context
Inclusion Criteria	Healthy, ambulatory, adult participants between 18 and 65 years of age, have accumulated less than 150 minutes of MVPA in the past week (assessed using the Active Australia Survey), and own or have access to a mobile phone compatible with both the Fitbit One and Jawbone UP
Exclusion Criteria	Unclear
Description of Study Protocol	Recruitment: Convenience sampling at three campuses of a large Australian metropolitan university via an email advertisement to staff that included study information and participant eligibility criteria Design: On two occasions for seven days each, participants wore an ActiGraph GT3X+ accelerometer on their right hip and either a hip-worn Fitbit One or wrist-worn Jawbone UP activity tracker. Daily estimates of steps and very active minutes were derived from the Fitbit One, and steps, active time, and longest idle time were derived from the Jawbone UP. Daily estimates of steps, MVPA, and longest sedentary bout were derived from the corresponding days of ActiGraph data. Blinding used (if applicable): Data were extracted via the users' accounts and entered into an Excel spreadsheet by a research assistant Intervention (if applicable): Data were collected as part of a larger, 12- week physical activity intervention study that included three groups that were randomly allocated to wear a Fitbit One, Jawbone UP, or standard pademeter. Demographia and anthronometria data were collected at

baseline. Data for this substudy were collected at mid- and postintervention when participants concurrently wore an ActiGraph GT3X+ accelerometer. On these two occasions for seven days each, participants wore an ActiGraph GT3X+ accelerometer on their right hip and either a hip-worn Fitbit One (n=14) or wrist-worn Jawbone UP (n=15) activity tracker. Participants were instructed to wear the devices during waking hours, removing them for water-based activities or contact sports, but were not required to keep wear logs in order to improve the free-living fidelity of the devices over the 12-week intervention. Participants were able to input activity sessions, such as swimming or contact sports, through the "log workout" function in the Jawbone UP app and the "track exercise" feature in the Fitbit app. Daily estimates of steps and very active minutes were derived from the Fitbit One, and steps, active time, and longest idle time were derived from the Jawbone UP. Daily estimates of steps, MVPA, and longest sedentary bout were derived from the corresponding days of ActiGraph data. Data were extracted via the users' accounts and entered into an Excel spreadsheet by a research assistant. Statistical Analysis: Descriptive statistics (n, mean, standard deviation, and prevalence) were calculated for demographic and physical measures. Absolute agreement was examined using intraclass correlation coefficients (ICC) and 95% confidence intervals. Correlation was assessed using Pearson correlation coefficient or Spearman rank correlation coefficient when data were non-normally distributed with 95% confidence intervals. Bland-Altman plots were used to examine the differences between all outcomes, with mean bias and 95% limits of agreement reported. Linear regression was used to examine whether mean difference and limits of agreement varied across mean values of Fitbit One or Jawbone UP and ActiGraph outcomes. Cohen's kappa statistic was used to assess the agreement between devices for classification of active versus inactive based first on achieving 10,000 steps or more per day (default step goal on both devices) and second on

	achieving 30 minutes per day or more of MVPA (comparable with public
	health guidelines). P values were based on two-sided tests and were
	considered statistically significant at P<.05. Post hoc power calculations
	determined that a sample size of N=289 daily comparisons would detect
	correlations as low as .17 with 80% power and 5% alpha.
	Timing of Measurements: Data was collected continuously as participants
	were instructed to wear either a Fitbit One or Jawbone UP activity tracker
	every day during the 12-week physical activity intervention. Daily
	estimates of steps, active time, and longest idle time were extracted from
	the users' accounts. Participants concurrently wore an ActiGraph GT3X+
	accelerometer at mid- and post-intervention for seven days. Daily
Data Collection	estimates of steps, MVPA, and longest sedentary bout were derived from
Summary	the corresponding days of ActiGraph data.
	Dependent Variables: Steps and very active minutes according to the
	Fitbit One; steps, active time, and longest idle time according to the
	Jawbone UP
	Independent Variables: Free living context (e.g. participants' daily
	obligations, lifestyles, level of physical fitness, stress levels)
	Control Variables: Steps, MVPA, and longest sedentary bout according to
	the ActiGraph GT3X+ accelerometer
	Initial: 32 ( Males Females)
	Attrition (final N): Only 29 participants provided valid data for the
	current analyses, 3 males and 26 females
Decemination of	Age: 39.6, SD 11.0 years
Actual Data	Ethnicity: Unclear
Sample	Other relevant demographics: 25 participants (86%) completed tertiary
	education
	Anthropometrics: Mean BMI: 25.9, SD: 5.0 kg/m2
	Location: Australia

	Key Findings: Correlations for steps and MVPA were strong for both
	devices, although higher for Fitbit One (r=.85 for steps and $\rho$ =.80 for
	MVPA) than for Jawbone UP (r=.75 for steps and $\rho$ =.75 for MVPA). The
	correlation between Jawbone UP longest idle time and ActiGraph longest
	sedentary bout was poor ( $\rho$ =.19). Absolute agreement (ICC) was
	acceptable for ActiGraph and Fitbit One steps (.90) and MVPA (.72) and
	Jawbone UP steps (.79). However, agreement was weak between
	ActiGraph and Jawbone UP estimates of MVPA (.56) and longest idle
	time (.08). For the estimation of steps, 95% limits of agreement were
	unbiased for both devices, although limits were wider for Jawbone UP
	than for Fitbit One (5290 and 3567 steps/day). When absolute values
	were calculated, both devices overestimated steps (Fitbit One: mean bias
	767, 95% limits of agreement –2800 to 4334; Jawbone UP: mean bias
	1178, 95% limits of agreement –4112 to 6468). For the estimation of
	MVPA, bias was evident for both the mean difference and the limits of
Summary of	agreement for both the Fitbit One and the Jawbone UP. When absolute
Results	values were calculated, the Fitbit One underestimated MVPA by a mean
	19.2 minutes/day (95% limits of agreement –39.2 to 5.5), whereas the
	Jawbone UP overestimated by a mean of 38.1 minutes/day (95% limits of
	agreement 5.8-65). For the estimation of longest sedentary bout, the
	limits of agreement were unbiased but wide (mean difference $\pm 88$
	minutes), varying by up to 150% of the mean estimate according to
	ActiGraph. Using the criterion of at least 10,000 steps per day, agreement
	between the Fitbit One and ActiGraph for the classification of active
	versus inactive was substantial ( $\kappa$ =.68, P<.001). The Fitbit One correctly
	classified 95% of days as active and 79% of days as inactive. Agreement
	between the Jawbone UP and ActiGraph was moderate ( $\kappa$ =.52, P<.001).
	The Jawbone UP correctly classified 90% of days as active and 80% of
	days as inactive. Using the criterion of at least 30 minutes/day of MVPA,
	agreement between the Fitbit One and ActiGraph was fair ( $\kappa$ =.39,
	P<.001). The Fitbit One correctly classified 40% of days as active and

	100% of days as inactive. Agreement between the Jawbone UP and
	ActiGraph was slight ( $\kappa$ =.14, P=.001). The Jawbone UP correctly
	classified 100% of days as active and 12% of days as inactive.
	Other Findings:
	The findings reported in this study suggest that both activity trackers have
	utility for counting steps in free-living settings, with both devices
	overestimating daily steps by only 5% to 15% compared with ActiGraph
	(Fitbit One: 8%; Jawbone UP: 14%). Both devices were less accurate
	measuring MVPA than steps, with correlations of .56 to .80 for both
	devices against ActiGraph data. Despite reasonable correlations, the
	Fitbit One underestimated MVPA by 46%, while the Jawbone UP
	overestimated MVPA by 50%. Findings indicate that the validity of the
	Jawbone UP measure of sedentary behavior (longest idle time) compared
Author	with ActiGraph-determined "longest sedentary bout" was poor. Both
Conclusion	devices accurately classified more than 80% of the sample days as active
	or inactive based on the 10,000 steps criterion; however, days were
	frequently misclassified for meeting public health guidelines of 30
	minutes/day of MVPA. Due to modest accuracy and systematic bias, both
	activity trackers are better suited as self-monitoring tools (e.g. for the
	public consumer or in behavior change interventions) rather than for
	evaluation of research outcomes. The outcomes that relate to sedentary
	behavior and MVPA should be used with caution for both consumers and
	researchers alike.
	Strengths: concurrent assessment of two popular brands of activity
	trackers on the market and two popular wear locations (wrist and waist),
Reviewer	large number of daily observations for comparison, free-living setting
Comments	which improves ecological validity and takes previous laboratory studies
	into a real-world setting, sample had good wear compliance, and the

	Limitations: predominantly female, healthy, middle-aged sample which
	limits the generalizability of the findings, and the study could not control
	for wear time of the activity trackers which may explain some of the large
	absolute differences between the devices and the ActiGraph
Funding Source	Start-Up Grant from The University of Queensland. Drs. Gomersall and
	Pavey were supported by an Australian National Health and Medical
	Research Council (NHMRC) program grant (NHMRC no: 569940).
	Authors declared no conflicts of interest.

## **Quality Criteria Checklist: Primary Research**

Symbols Used	Explanation
	<b>Positive</b> – Indicates that the report has clearly addressed issues of
+	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
0	exceptionally week

Select a rating from the drop-down menu ↓

R	elevance Questions		
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

V۶	alidity Questions		
1.	Was the <u>research question</u> clearly stated?	1	Yes
	1.1. Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	Yes
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3. Were the target population and setting specified?	1.3	Yes
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Unclear
	progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Unclear
	2.2. Were criteria applied equally to all study groups?	2.2	Yes
	2.3. Were health, demographics, and other characteristics of subjects described?	2.3	Yes

	2.4.	Were the subjects/patients a representative sample of the relevant	2.4	Unclear
		population?	2.1	
3.	<b>Were</b> 3.1.	study groups comparable? Was the method of assigning subjects/patients to groups described and	3	Unclear
	3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	Unclear
	3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	3.2	Unclear
	3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.3	Yes
	3.5.	appropriate adjustments in statistical analysis? If case control study, were potential confounding factors comparable for cases	3.4	N/A
		criterion is not applicable. Criterion may not be applicable in some cross- sectional studies.)	3.5	N/A
	3.6.	If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	Yes
4.	Was n	nethod of handling <u>withdrawals</u> described?	4	N/A
	4.1.	Were follow up methods described and the same for all groups? Was the number, characteristics of withdrawals (i.e., dronouts, lost to follow,	4.1	N/A
	4.2.	up, attrition rate) and/or response rate (cross-sectional studies) described for	12	N/A
		each group? (Follow up goal for a strong study is 80%.)	4.2	IN/A
	4.3.	Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	No
	4.4. 4 5	Were reasons for withdrawais similar across groups	4.4	N/A
	4.5.	results of test under study?	4.5	Yes
5.	Was b	linding used to prevent introduction of bias?	-	V
	5.1.	In intervention study, were subjects, clinicians/practitioners, and investigators	5	Yes
	5.2.	blinded to treatment group, as appropriate? Were data collectors blinded for outcomes assessment? (If outcome is	5.1	N/A
		measured using an objective test, such as a lab value, this criterion is assumed to be met.)	5.2	Yes
	5.3.	In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?	5.3	N/A
	5.4.	In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4	N/A
	5.5.	In diagnostic study, were test results blinded to patient history and other test results?	5.5	Yes
6.	Were	intervention/therapeutic regimens/exposure factor or procedure and any	6	Yes
	6.1.	In RCT or other intervention trial, were protocols described for all regimens	6.1	N/A
		studied?	( )	N/
	6.2.	In observational study, were interventions, study settings, and	6.2	Yes
	6.3.	clinicians/provider described? Was the intensity and duration of the intervention or exposure factor sufficient	6.3	Yes
		to produce a meaningful effect?	6.4	Yes
	o.4.	was the amount of exposure and, if relevant, subject/patient compliance measured?	6.5	Yes
	6.5.	Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.6	N/A

	6.6. Were extra or unplanned treatments described?	6.7	Yes				
	6.7. Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?						
	6.8. In diagnostic study, were details of test administration and replication sufficient?	6.8	Yes				
7.	Were <u>outcomes</u> clearly defined and the <u>measurements valid and reliable</u> ?	7	Yes				
	7.1. Were primary and secondary endpoints described and relevant to the	7.1	Yes				
	question ?	7.2	N/A				
	7.3 Was the period of follow-up long enough for important outcome(s) to occur?	7.2					
	7.4 Were the observations and measurements based on standard valid and	7.3	N/A				
	reliable data collection instruments/tests/procedures?	7.4	Yes				
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes				
	7.6. Were other factors accounted for (measured) that could affect outcomes?	76	No				
	7.7. Were the measurements conducted consistently across groups?	7.0	Vaa				
_		1.1	res				
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome	8	Yes				
	8.1 Were statistical analyses adequately described the results reported	8.1	Vec				
	appropriately?	0.1	105				
	8.2. Were correct statistical tests used and assumptions of test not violated?	8.2	Yes				
	8.3. Were statistics reported with levels of significance and/or confidence intervals?	83	Ves				
	8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?		105				
			N/A				
	8.5. Were adequate adjustments made for effects of confounding factors that	8.5	N/A				
	might have affected the outcomes (e.g., multivariate analyses)?	86	Ves				
	8.6. Was clinical significance as well as statistical significance reported?		100				
	8.7. If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A				
9.	Are <u>conclusions supported by results</u> with biases and limitations taken into	9	Yes				
	consideration?	9.1	Yes				
	9.1. Is there a discussion of findings?	9.2	Yes				
10	s.z. Are blases and study initiations identified and discussed?	10	Vac				
10.	10.1 Were sources of funding and investigators' affiliations described?	10 1	I ts				
	10.2. Was there no apparent conflict of interest?	10.1	I ES				
М		10.2	1 65				
If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus							

(-) symbol on the Evidence Worksheet.

NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

	Contrinsi I. Description of the Division of (2016) Constraints Marchine Astro-
Citation: write it in AMA format	Gualtieri, L., Rosenbluth, S., & Phillips, J. (2016). Can a Free wearable Activity
as jound in JADA.	Tracker Change Benavior? The Impact of Trackers on Adults in a Physician-Led
	Wellness Group. JMIR Research Protocols, 5(4), e237. doi:10.2196/resprot.6534
Study design: Use algorithm –	Non-randomized crossover trial
RC1, conort, etc	
Study Class (A,B,C,D)	C
Research Quality Rating	NEUTRAL (Ø)
This rating tells if the research	
design is good (+), bad (-) or	
neutral ( $\varnothing$ )	
This is determined by the quality	
criteria list. Delete the ratings	
that do not apply (i.e. if positive,	
delete minus/negative and	
neutral).	
	Purpose/Population Studied/Practice Studied
Research purpose: What is the	To investigate the use of wearable activity trackers by adults with chronic medical
research question being	conditions who have never used trackers previously. Specifically, the researchers
investigated in the study?)	aimed to determine (1) if participants would accept and use trackers to increase
	their physical activity; (2) if there were barriers to use besides cost and training; (3)
	if trackers would educate participants on their baseline and ongoing activity levels
	and support behavior change; and (4) if clinical outcomes would show
	improvements in participants' health.
Inclusion criteria: requirements	Patients had to be part of the private practice (Family Doctors, LLC), have at least
for study eligibility	one chronic medical condition, and be over 18 years of age
Exclusion criteria (conditions	Patients who could not comprehend and speak English, or if they had advanced
that make individual ineligible)	dementia
Recruitment	Through the Family Doctors, LLC Facebook page, brochures in the office, word of
	mouth, and informal mentions from staff that patients would receive a free
	wearable activity tracker. Cost to patients was a US \$150 program fee, plus
Blinding used: some of the	NA
billing used: some of the	NA
from knowing cartain information	
that might lead to conscious or	
unconscious bias on their part	
invalidating the results	
Description of study protocol	This study was conducted with patients (N=10) in a 12-week physician-led
What happened in the study?	wellness group offered by Family Doctors, LLC. Patients were given Withings
· · · · · · · · · · · · · · · · · · ·	Pulse wearable activity trackers in the second week of the wellness group and were
	interviewed two to four weeks after it ended. Study investigators analyzed the
	interview notes to extract themes about the participants' attitudes and behavior
	changes and collected and analyzed participants' clinical data over the course of the
	study.
Intervention: Describe	The wellness group was designed as a 12-week program with two-hour meetings
interventions, regimens, risk	every week, during which patients received guidance and teaching from health
factors, or procedures studied.	experts on physical activity, nutrition, mental health, mindfulness, and sleep. At
_	week two of the 12-week wellness group, all participants were given a new
	Withings Pulse, a wearable activity tracker that measures step count, calories
	burned, distance walked, heart rate, and sleep. Participants were given instructions
	developed by the research team on the setup and use of the activity tracker.
	Researchers assisted seven participants in setting up their devices, while the

	remaining three felt confident in setting up their devices independently. Participants were given guidance on how to select their daily step count goal. Some used the default step goal of 10,000 steps per day, while those with significant physical limitations used a goal personalized to their needs, with instructions to slowly increase their daily and weekly step count as their health permitted. In alignment with the philosophy of the wellness group, the use of activity trackers was discussed with participants as a way to build better health habits and create lifestyle change. Researchers helped troubleshoot or answer participant questions about the activity trackers during weekly meetings, by phone, and by email. All 10 patients who completed the 12-week program participated in semi-structured phone interviews, consisting of 18 open-ended questions with potential follow-up statements to encourage further responsiveness, which occurred at weeks 14, 15, and 16. Researchers recorded age, systolic blood pressure (SBP), diastolic blood pressure (DBP), low-density lipoprotein (LDL), and body weight at the start and end of the intervention.
Statistical analysis: List tests,	Thematic analyses were conducted through reviews of interview notes to identify
significance level set a priori	underlying themes in participant experiences. Transcripts were manually reviewed
$(\alpha=0.05; include intent to treat$	tor common language and word choice, followed by multiple discussion sessions to
analysis if applicable; note if	determine significance and prevalence of themes. Paired t tests and P values were calculated and P values lass than 05 were causidered as significant.
Timing of monsurements, where	calculated, and r values less than .05 were considered as significant.
outcomes were measured; usually baseline and one or more later times	Researchers recorded age, SBP, DBP, LDL, and body weight at the start and end of the 12-week intervention. Step count, calories burned, distance walked, heart rate, and sleep data was collected continuously throughout 11 weeks of the wellness program while participants wore their Withings Pulse activity tracker. Phone interviews occurred at weeks 14, 15, and 16, and lasted approximately 30 minutes each.
Dependent variables: outcomes that are measured or registered; variable whose change or different states the researcher wants to understand, explain, or predict	The amount of use, acceptance, and barriers associated with the Withings Pulse. Changes in step count, calories burned, distance walked, heart rate, and sleep data. Levels of physical activity and measurements of SBP, DBP, LDL, and body weight after the 12-week intervention.
Independent variables	Two-hour meetings every week, during which patients received guidance and
(intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying	teaching from health experts on physical activity, nutrition, mental health, mindfulness, and sleep
to understand)	Description levels of always of a stight, and many structure (ODD, DDD, IDI, 1)
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	baseline levels of physical activity and measurements of SBP, DBP, LDL, and body weight
<b>Initial n</b> (e.g. 731 (298 males,	11 participants
433 females)) Record number that entered	
<b>Final n</b> (attrition)	10 participants 2 males and 8 females
number of subjects that completed study	To participants, 2 maios and 6 females
Age usually mean or range	39 to 77 years
Ethnicity (if given)	Unclear

Other relevant demographics:	Primarily lower-income patients five (50%) worked full-time one (10%) worked
demographics describe the	part-time and four (40%) were retired. All patients suffered from at least one of
nonulation (students athletes etc)	the following chronic medical problems: overweight or obesity hypertension type
population (statents, attictes, etc)	2 diabetes hyperlinidemia and joint nain. All natients stated that they were first-
	time activity tracker users at the onset of the group
Anthronomotrioge a gruppe	All but and of the notion to use a supervisible or chase. Deceling lowels of physical
Anthropometrics: e.g. were	All but one of the patients was overweight of obese. Baseline levels of physical
groups same or all ferent on	activity, as assessed through patient interviews and group counseling, ranged from
important physical measures	almost entirely sedentary to moderately active.
(BMI, fitness level)	
Location: Where did the study	Family Doctors, LLC, a private practice in a suburban community north of Boston,
take place? City or country	Massachusetts
Summary of Results: Abstract	Over the 11 weeks of activity tracker use, improvements were seen in clinical
results including <i>quantitative data</i>	outcomes, physical activity behaviors, and attitudes towards the Withings Pulse.
and statistics. Include statistical	Participants lost an average of 0.5 pounds per week (SD 0.4), with a mean total
significance: P-values, confidence	weight loss of 5.97 pounds ( $P$ =.004). Other short-term clinical outcomes included
intervals (CI), relative risk (RR),	a 9.2% decrease in LDL levels ( $P$ =.038). Changes in blood pressure were non-
odds ratios (OR), likelihood ratio,	significant. All participants reported an increase in well-being, health education,
number needed to treat, power	physical activity, and confidence in their ability to lead more active lives.
analysis if available.	Researchers identified the following six major themes from the qualitative analysis
	of the post-intervention interview notes: (1) barriers to activity tracker purchase
	included cost, perceived value, and choice confusion; (2) attitudes towards the
	activity trackers shifted for many, from half of the participants expressing
	excitement and hope and half expressing hesitation or trepidation, to all participants
	feeling positive towards their tracker at the time of the interviews: (3) activity
	trackers served as educational tools for baseline activity levels: (4) activity trackers
	provided concrete feedback on physical activity which motivated behavior change.
	(5) activity tracker use reinforced wellness group activities and goals: and (6)
	although commitment to activity tracker use did not waver, external circumstances
	and dight communent to activity tracker use and not waver, external enclansances
	influenced some participants' ongoing use
	influenced some participants' ongoing use.
Author conclusion: paraphrase	influenced some participants' ongoing use. <i>Author's Conclusions</i> Findings suggest that adding activity trackers to wellness groups comprising
Author conclusion: paraphrase	influenced some participants' ongoing use. <i>Author's Conclusions</i> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and
<b>Author conclusion:</b> paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and
<b>Author conclusion:</b> paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and abaiaa confusion, which were removed by providing participants with free activity
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The harriers to purchase included cost, with free activity
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and opening support. Overall, this study domentated the advectional
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational hereafte to individue a charging their baseling activity levels, the increased activity
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self-
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost-
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use.Author's ConclusionsFindings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care,
Author conclusion: paraphrase that stated by study author in body of the report or abstract	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop.
Author conclusion: paraphrase         that stated by study author in         body of the report or abstract             Reviewer comments: Note	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers,
Author conclusion: paraphrase         that stated by study author in         body of the report or abstract         Best construct         Reviewer comments: Note         strengths and limitations of study;	influenced some participants' ongoing use.Author's ConclusionsFindings suggest that adding activity trackers to wellness groups comprisingprimarily older adults with chronic medical conditions can support education andbehavior change to be more physically active. Barriers need to be identified andremoved. In this study, the barriers to purchase included cost, perceived value, andchoice confusion, which were removed by providing participants with free activitytrackers. The barriers to use were removed by providing participants with initialtraining and ongoing support. Overall, this study demonstrated the educationalbenefits to individuals of learning their baseline activity levels, the increased self-efficacy arising from concrete feedback on physical activity that motivatedbehavior change, the positive attitudes that developed towards activity trackers, andimprovements in clinical outcomes. Findings also suggest that it may be cost-effective for physicians and other health care providers to provide free or heavilysubsidized trackers, along with training and support, to their patients, especiallythose who may most benefit from increasing their physical activity. A US \$60activity tracker that lowers the risk of chronic conditions by facilitating changes inhealth behaviors would be greatly beneficial compared to the health care,medication, or intervention costs required to treat illnesses after they develop.Strengths: providing
Author conclusion: paraphrase that stated by study author in body of the report or abstract         Beviewer comments: Note strengths and limitations of study; identify concerns that affect study	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers, initial training, and ongoing support to minimize barriers, and the multifaceted approach to health and wellness, which encouraged participants to incorporate
Author conclusion: paraphrase that stated by study author in body of the report or abstract         Beviewer comments: Note strengths and limitations of study; identify concerns that affect study validity and generalizability—	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers, initial training, and ongoing support to minimize barriers, and the multifaceted approach to health and wellness, which encouraged participants to incorporate gradual, evidence-based changes into their lives, promoting true lifestyle change
Author conclusion: paraphrase that stated by study author in body of the report or abstract         Beviewer comments: Note strengths and limitations of study; identify concerns that affect study validity and generalizability— your comments should be	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers, initial training, and ongoing support to minimize barriers, and the multifaceted approach to health and wellness, which encouraged participants to incorporate gradual, evidence-based changes into their lives, promoting true lifestyle change rather than "dieting" or being on an "exercise program"
Author conclusion: paraphrase that stated by study author in body of the report or abstract         Beviewer comments: Note strengths and limitations of study; identify concerns that affect study validity and generalizability— your comments should be italicized)	influenced some participants' ongoing use. Author's Conclusions Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers, initial training, and ongoing support to minimize barriers, and the multifaceted approach to health and wellness, which encouraged participants to incorporate gradual, evidence-based changes into their lives, promoting true lifestyle change rather than "dieting" or being on an "exercise program"
Author conclusion: paraphrase that stated by study author in body of the report or abstract         Best and the report of abstract         Reviewer comments: Note strengths and limitations of study; identify concerns that affect study validity and generalizability—your comments should be italicized)	influenced some participants' ongoing use. <u>Author's Conclusions</u> Findings suggest that adding activity trackers to wellness groups comprising primarily older adults with chronic medical conditions can support education and behavior change to be more physically active. Barriers need to be identified and removed. In this study, the barriers to purchase included cost, perceived value, and choice confusion, which were removed by providing participants with free activity trackers. The barriers to use were removed by providing participants with initial training and ongoing support. Overall, this study demonstrated the educational benefits to individuals of learning their baseline activity levels, the increased self- efficacy arising from concrete feedback on physical activity that motivated behavior change, the positive attitudes that developed towards activity trackers, and improvements in clinical outcomes. Findings also suggest that it may be cost- effective for physicians and other health care providers to provide free or heavily subsidized trackers, along with training and support, to their patients, especially those who may most benefit from increasing their physical activity. A US \$60 activity tracker that lowers the risk of chronic conditions by facilitating changes in health behaviors would be greatly beneficial compared to the health care, medication, or intervention costs required to treat illnesses after they develop. Strengths: providing all participants with free Withings Pulse activity trackers, initial training, and ongoing support to minimize barriers, and the multifaceted approach to health and wellness, which encouraged participants to incorporate gradual, evidence-based changes into their lives, promoting true lifestyle change rather than "dieting" or being on an "exercise program" Limitations: <i>small sample size, predominantly female sample, study design lacked a</i> <i>control group, blinding was not utilized</i> , and the results cannot separate the impact

Funding source: sponsorship was provided by Withings, who donated trackers to
RecycleHealth, a non-profit at Tufts University, who provided them to participants
in the Family Doctors, LLC Wellness Group. The authors declared no conflicts of
interest.

Table 3.2.a. Quality Criteria Checklist: Primary Research

DELEWANCE OUESTIONS					
<b>RELEVANCE QUESTIONS</b> <b>Citation:</b> <i>write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)</i> Gualtieri, L., Rosenbluth, S., & Phillips, J. (2016). Can a Free Wearable Activity Tracker Change Behavior? The Impact of Trackers on Adults in a Physician-Led Wellness Group.		Y E S	N O	U N C L E	N A
JMIR Research Protocols, 5(4), e237. doi:10.2196/resprot.6534				AR	
1. Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some epidemiological studies)	1	X			
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/ population group would care about?	2	X			
3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	X			
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "yes", the report is eligible for design the Evidence Quality Worksheet, depending on answers to the following validity questions.	ation w	vith a	plu	s (+	) on
VALIDITY QUESTIONS		v		NI	IN
<b>1.</b> Was the <u>research question</u> clearly stated? This is usually stated at end of the introduction and just before methods section.		ES		0 1	N A
		У	K	1	A A
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified? <i>This is often called the treatment and explained in the methods section.</i>	1.1	Σ	K .		
<ul> <li>1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?</li> <li>These are sometimes called the endpoints; the results section reports the outcomes, but this information should be in the methods section, too</li> </ul>	1.2	У	K		
1.3 Were the target population and setting specified? <i>The target population is group for whom findings may be applicable; look for this in the</i> <i>introduction and in the methods section</i>	1.3	У	K		
2. Was the <u>selection</u> of study subjects/patients free from bias?		Y E S	]		U N N A C C X
<ul> <li>2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?</li> <li>The authors should give several points about the inclusion/exclusion criteria such as the age range of the subjects, disease condition (like hyperlipidemia) required for inclusion. Exclusion criteria should be listed, too, although some are understood. For example if the ages for inclusion are 18 to 70, the authors will probably not specifically note that children and people over age 70 were excluded. Most of the time, however, subjects may be excluded for certain characteristics such as being pregnant or having some disease (like CHD).</li> </ul>	2.1	>			
2.2 Were criteria applied equally to all study groups?	2.2				X
<ul> <li>2.3 Were health, demographics, and other characteristics of subjects described? There is usually a Table 1 summarizing demographics and characteristics at baseline. Groups are not different if the P-Value is &gt; 0.05. If there has been a previous paper describing the study population, that paper may be referenced and you would need to go back to the original publication to see that Table 1.</li> <li>2.4 Were the subjects/patients in a representative sample of the relevant population?</li> </ul>	2.3	>			x
2.4 were the subjects/patients in a representative sample of the relevant population?	2.4			-	1

	The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion					
	criteria.					
3. \	Were <u>study groups comparable</u> ?		Y E	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		s	-	C	
	Groups are <u>not</u> different if the P-Value is $> 0.05$ .				E	
					A R	Х
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1		Х		
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
	demographics) similar across study groups at baseline? See Table I for this - there					
	should be no significant differences across study groups in an intervention study.					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	X			
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
	prior study as a control group; that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control.	2.4				v
	5.4 If conort study or cross-sectional study, were groups comparable on important	5.4				Λ
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross sectional study should not be different from each other:					
	if they are a strong study will utilize statistical techniques such as multivariate					
	if they are, a strong study will utilize statistical techniques such as multivariate analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections					
	3.5. If case control study, were potential confounding factors comparable for cases and	35				X
	controls? If case series or trial with subjects serving as own control this criterion is	5.5				11
	not applicable. Criterion may not be applicable in some cross-sectional studies					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6				Х
	reference standard (e.g. "gold standard")?					
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4.	Was method of handling <u>withdrawals</u> described?		Y E	N O	U N	N A
			s		C L	
					E	
			Х		R	
	4.1 Were follow up methods described and the same for all groups?	4.1				Х
	4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up,	4.2	Х			
	attrition rate) and/or response rate (cross-sectional studies) described for each group?		1			
1	(Follow up goal for a strong study is 80 %.)					
1	This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
1	note that on the worksheet (as a note in the results section or in the reviewer					
	<i>comments at the very Dottom)</i>	1.2		v		
	4.5 were all enioned subjects/patients (in the original sample) accounted for?	4.3	1	A		
	N) # remaining at end of study period (final N) Sometimes the reasons that subjects		1			
	withdrew or were dronned is given in the figure or in the text (results section)		1			
	minuter of were dropped is given in the figure of in the text (results section).	1		1		

<b></b>		4.4	1	П		V
	4.4 Were reasons for withdrawals similar across groups?	4.4				Х
	If there is a large attrition from one group and not others, you would want to look for a reason why: the answer to this question would then be no					
	4.5. If diagnostic test, was decision to perform reference test not dependent on results of	45				X
	test under study?	ч.5				1
	The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					
	BMI was > 35 but bioimpedance analyzer indicated body fat < 30%.					
5.	Was blinding used to prevent introduction of bias?		Y	N	UN	N
			S	0	C	А
					L E	
				Х	A R	
	5.1. In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1		v		
	to treatment group as appropriate?	5.1		Λ		
	The key term is as appropriate For example in the Lim et al 2008 study the investigators					
	studied the effect of MNT on linid levels in hypercholesterolemic nations. It was					
	an RCT but obviously the subjects and practitioners knew who was getting MNT					
	and who was not Therefore you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2		X		
	using an objective test, such as a lab value, this criterion is assumed to be met.)					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).	L				
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	influenced by exposure status?					
	Establish who is a case and who is a control at the beginning of the study.					V
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5	v	N	П	X
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any		E	0	N	A
	comparison(s) described in detail? were intervening factors described?		5		L	
			v		E A	
		(1			R	
	6.1 In RC1 or other intervention trial, were protocols described for all regimens studied?	6.1	X			
	6.2 In observational study, were interventions, study settings, and clinicians/provider described?	6.2				Х
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
	produce a meaningful effect?					
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
	difference in lab values for cholesterol; however, 12 days would not be long					
	enough)					
	6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured? How long did the treatment last? Did the patient follow directions?	6.4	Х			
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5	Х			
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)	ļ				
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the		1			
	discussion section. It is likely there were no unplanned treatments, so a "no"		1			
	answer is not a problem overall.					

	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups? <i>For a study to be valid and unbiased, it is important that this be yes.</i>	6.7				Х
	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8				Х
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E A R X	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	Х			
	7.2 Were nutrition measures appropriate to question and outcomes of concern? Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul><li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li><li>Check that surveys were validated.</li></ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability.</i>	7.5			Х	
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6			X	
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A R X	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1			Х	
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals ( <i>mean</i> $\pm$ <i>CI</i> )	8.3	Х			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4		X		
	8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?	8.5		Х		

Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a					
smoking, etc). Assumes data is valia and reduces a targer number of variables to a smaller number .lust answer ves or no that multivariate analyses were used					
8.6 Was clinical significance as well as statistical significance reported?	8.6	X			
Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was					
reduced from 229.2 $\pm$ 158 to 181.3 $\pm$ 16.3 (P<0.001); This includes: statistical					
significance (P-value) and clinical significance (compare to standard of $< 200$					
mg/do for normal cholesterol). A problem can occur when only statistical					
significance is reported. Reducing cholesterol from 300 to 250 might be statistically					
significant, but clinically it is still abnormal.					
8.7 If negative findings, was a power calculation reported to address type 2 error?	8.7				Х
Type II ( $\beta$ error is a false negative that happens when the investigators fail to reject the					
<u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say					
something like "a sample size of $n=xx$ is needed to provide 80% power."					
9. Are <u>conclusions</u> supported by results with biases and limitations taken into		Y E	N O	U N	N A
consideration?		s		C L	
		v		E A	
		X		R	
9.1 Is there a discussion of findings?	9.1	Х			
Answer yes or no.					
9.2 Are biases and study limitations identified and discussed?	9.2	Х			
This will be in the discussion of finding section that follows the results		v	N	п	N
10. Is bias due to study's <u>funding or sponsorship</u> unlikely?		E	0	N	A
Be carejui nere – ij <u>bias</u> is <u>uniikely</u> , answer TES.		5		L	
		v		E A	
	10.1	Λ		R	
10.1 Were sources of funding and investigators' affiliations described?	10.1	Х			
• Look just under the abstract, or					
• The funding may be acknowledged at the end of the paper					
• Just because the work was funded by industry does not mean the study was biased.	10.2	v			
10.2 was there no apparent conflict of interest?	10.2	Х			
If an investigator is testing a piece of equipment, process or aring that s/ne developed, it					
SVMDOI		1		1	
STMDOL					
MINUS/NEGATIVE (-)					
If most (six or more) of the answers to the above validity questions are "no," the report should a	be design	ated	with	a	
minus (-) symbol on the Evidence Quality Worksheet.					
NEUTRAL (Ø)					
If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is except	ionally st	rong	the		
report should be designated with a neutral (ø) symbol on the Evidence Quality Worksheet.					
PLUS/POSITIVE (+)					

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

right. Academy of Nutrition and Dietetics

Academy of Nutrition and Dietetics Evidence Analysis Library® Worksheet Template and Quality Criteria Checklist: Primary Research

Citation	Gualtieri, L., Rosenbluth, S., & Phillips, J. (2016). Can a Free Wearable Activity Tracker Change Behavior? The Impact of Trackers on Adults in a Physician-Led Wellness Group. JMIR Research Protocols, 5(4), e237. doi:10.2196/resprot.6534
Study Design	Non-randomized crossover trial
Class	C
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\boxtimes \otimes$ (Neutral)
Research Purpose	To investigate the use of wearable activity trackers by adults with chronic medical conditions who have never used trackers previously. Specifically, the researchers aimed to determine (1) if participants would accept and use trackers to increase their physical activity; (2) if there were barriers to use besides cost and training; (3) if trackers would educate participants on their baseline and ongoing activity levels and support behavior change; and (4) if clinical outcomes would show improvements in participants' health.
Inclusion Criteria	Patients had to be part of the private practice (Family Doctors, LLC), have at least one chronic medical condition, and be over 18 years of age
Exclusion Criteria	Patients who could not comprehend and speak English, or if they had advanced dementia
Description of Study Protocol	brochures in the office, word of mouth, and informal mentions from staff that patients would receive a free wearable activity tracker. Cost to patients was a US \$150 program fee, plus insurance co-payments. Design: This study was conducted with patients (N=10) in a 12-week physician-led wellness group offered by Family Doctors, LLC. Patients were given Withings Pulse wearable activity trackers in the second week of the wellness group and were interviewed two to four weeks after it ended. Study investigators analyzed the interview notes to extract themes about the participants' attitudes and behavior changes and collected and analyzed participants' clinical data over the course of the study. Blinding used (if applicable): NA Intervention (if applicable): The wellness group was designed as a 12- week program with two-hour meetings every week, during which patients received guidance and teaching from health experts on physical activity,

	nutrition, mental health, mindfulness, and sleep. At week two of the 12-
	week wellness group, all participants were given a new Withings Pulse, a
	wearable activity tracker that measures step count, calories burned,
	distance walked, heart rate, and sleep. Participants were given
	instructions developed by the research team on the setup and use of the
	activity tracker. Researchers assisted seven participants in setting up their
	devices, while the remaining three felt confident in setting up their
	devices independently. Participants were given guidance on how to select
	their daily step count goal. Some used the default step goal of 10,000
	steps per day, while those with significant physical limitations used a goal
	personalized to their needs, with instructions to slowly increase their daily
	and weekly step count as their health permitted. In alignment with the
	philosophy of the wellness group, the use of activity trackers was
	discussed with participants as a way to build better health habits and
	create lifestyle change. Researchers helped troubleshoot or answer
	participant questions about the activity trackers during weekly meetings,
	by phone, and by email. All 10 patients who completed the 12-week
	program participated in semi-structured phone interviews, consisting of
	18 open-ended questions with potential follow-up statements to encourage
	further responsiveness, which occurred at weeks 14, 15, and 16.
	Researchers recorded age, systolic blood pressure (SBP), diastolic blood
	pressure (DBP), low-density lipoprotein (LDL), and body weight at the
	start and end of the intervention.
	Statistical Analysis: Thematic analyses were conducted through reviews
	of interview notes to identify underlying themes in participant
	experiences. Transcripts were manually reviewed for common language
	and word choice, followed by multiple discussion sessions to determine
	significance and prevalence of themes. Paired t tests and P values were
	calculated, and P values less than .05 were considered as significant.
Data Collection Summary	

	Timing of Measurements: Researchers recorded age, SBP, DBP, LDL,
	and body weight at the start and end of the 12-week intervention. Step
	count, calories burned, distance walked, heart rate, and sleep data was
	collected continuously throughout 11 weeks of the wellness program
	while participants wore their Withings Pulse activity tracker. Phone
	interviews occurred at weeks 14, 15, and 16, and lasted approximately 30
	minutes each.
	Dependent Variables: The amount of use, acceptance, and barriers
	associated with the Withings Pulse. Changes in step count, calories
	burned, distance walked, heart rate, and sleep data. Levels of physical
	activity and measurements of SBP, DBP, LDL, and body weight after the
	12-week intervention.
	Independent Variables: Two-hour meetings every week, during which
	patients received guidance and teaching from health experts on physical
	activity, nutrition, mental health, mindfulness, and sleep
	Control Variables: Baseline levels of physical activity and measurements
	of SBP, DBP, LDL, and body weight
	Initial: 11 ( Males Females)
	Attrition (final N): 10 (2 males and 8 females)
	Age: 39 to 77 years
	Ethnicity: Unclear
	Other relevant demographics: Primarily lower-income patients, five
	(50%) worked full-time, one (10%) worked part-time, and four (40%)
Description of	were retired. All patients suffered from at least one of the following
Actual Data Sample	chronic medical problems: overweight or obesity, hypertension, type 2
1	diabetes, hyperlipidemia, and joint pain. All patients stated that they were
	first-time activity tracker users at the onset of the group.
	Anthropometrics: All but one of the patients was overweight or obese.
	Baseline levels of physical activity, as assessed through patient interviews
	and group counseling, ranged from almost entirely sedentary to
	moderately active.

	Location: Family Doctors, LLC, a private practice in a suburban
	community north of Boston, Massachusetts
	Key Findings: Over the 11 weeks of activity tracker use, improvements
	were seen in clinical outcomes, physical activity behaviors, and attitudes
	towards the Withings Pulse. Participants lost an average of 0.5 pounds
	per week (SD 0.4), with a mean total weight loss of $5.97$ pounds (P=.004).
	Other short-term clinical outcomes included a 9.2% decrease in LDL
	levels (P=.038). Changes in blood pressure were non-significant. All
	participants reported an increase in well-being, health education, physical
	activity, and confidence in their ability to lead more active lives.
	Researchers identified the following six major themes from the qualitative
	analysis of the post-intervention interview notes: (1) barriers to activity
Summary of	tracker purchase included cost, perceived value, and choice confusion; (2)
Results	attitudes towards the activity trackers shifted for many, from half of the
	participants expressing excitement and hope and half expressing
	hesitation or trepidation, to all participants feeling positive towards their
	tracker at the time of the interviews; (3) activity trackers served as
	educational tools for baseline activity levels; (4) activity trackers provided
	concrete feedback on physical activity, which motivated behavior change;
	(5) activity tracker use reinforced wellness group activities and goals; and
	(6) although commitment to activity tracker use did not waver, external
	circumstances influenced some participants' ongoing use.
	Other Findings:
	Findings suggest that adding activity trackers to wellness groups
	comprising primarily older adults with chronic medical conditions can
	support education and behavior change to be more physically active.
Author	Barriers need to be identified and removed. In this study, the barriers to
Conclusion	purchase included cost, perceived value, and choice confusion, which
	were removed by providing participants with free activity trackers. The
	barriers to use were removed by providing participants with initial

	training and ongoing support. Overall, this study demonstrated the
	educational benefits to individuals of learning their baseline activity
	levels, the increased self-efficacy arising from concrete feedback on
	physical activity that motivated behavior change, the positive attitudes
	that developed towards activity trackers, and improvements in clinical
	outcomes. Findings also suggest that it may be cost-effective for
	physicians and other health care providers to provide free or heavily
	subsidized trackers, along with training and support, to their patients,
	especially those who may most benefit from increasing their physical
	activity. A US \$60 activity tracker that lowers the risk of chronic
	conditions by facilitating changes in health behaviors would be greatly
	beneficial compared to the health care, medication, or intervention costs
	required to treat illnesses after they develop.
	Strengths: providing all participants with free Withings Pulse activity
	trackers, initial training, and ongoing support to minimize barriers, and
	the multifaceted approach to health and wellness, which encouraged
	participants to incorporate gradual, evidence-based changes into their
Reviewer	lives, promoting true lifestyle change rather than "dieting" or being on
Comments	an "exercise program"
	Limitations: Limitations: small sample size, predominantly female
	sample, study design lacked a control group, blinding was not utilized,
	and the results cannot separate the impact of the wellness group
	education and support from that of the activity tracker use
	Sponsorship was provided by Withings, who donated trackers to
	RecycleHealth, a non-profit at Tufts University, who provided them to
Funding Source	participants in the Family Doctors, LLC Wellness Group. The authors
	declared no conflicts of interest.

## Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
4	<b>Positive</b> – Indicates that the report has clearly addressed issues of
÷	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.

exceptionally week
--------------------

		Select a dro	rating from the p-down menu ↓		
Re	Relevance Questions				
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Yes		
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes		
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	Yes		
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes		

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Va	alidity	Questions		
1.	Was tl	ne <u>research question</u> clearly stated?	1	Yes
	1.1.	Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes
	12	identified? Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3.	Were the target population and setting specified?	1.3	Yes
2.	Was t	ne <u>selection</u> of study subjects/patients free from bias?	2	Unclear
	2.1.	Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes
	2.2.	without omitting criteria critical to the study? Were criteria applied equally to all study groups?	2.2	N/A
	2.3. 2.4.	Were health, demographics, and other characteristics of subjects described? Were the subjects/patients a representative sample of the relevant	2.3	Yes
		population?	2.4	Unclear
3.	<b>Were</b> 3.1.	study groups comparable? Was the method of assigning subjects/patients to groups described and	3	N/A
		unbiased? (Method of randomization identified if RCT)		
	3.2.	Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	No
	3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	3.2	N/A
	3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.3	Yes
	3.5.	appropriate adjustments in statistical analysis? If case control study, were potential confounding factors comparable for cases	3.4	N/A
		criterion is not applicable. Criterion may not be applicable in some cross- sectional studies.)	3.5	N/A
	3.6.	If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	N/A
4.	Was m	ethod of handling withdrawals described?	4	Yes

	4.1. Were follow up methods described and the same for all groups?	4.1	N/A
	4.2. Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow	4.0	X7
	up, attrition rate) and/or response rate (cross-sectional studies) described for	4.2	Yes
	each group? (Follow up goal for a strong study is 80%.)	4.3	No
	4.3. Were reasons for withdrawals similar across groups	4.4	N/A
	4.5. If diagnostic test, was decision to perform reference test not dependent on		
	results of test under study?	4.5	N/A
5	Was <u>blinding</u> used to prevent introduction of bias?	5	No
	5.1. In intervention study, were subjects, clinicians/practitioners, and investigators	5	110
	blinded to treatment group, as appropriate?	5.1	No
	5.2. Were data collectors blinded for outcomes assessment? (If outcome is		
	to he met )	5.2	No
	5.3. In cohort study or cross-sectional study, were measurements of outcomes and		
	risk factors blinded?	5.3	N/A
	5.4. In case control study, was case definition explicit and case ascertainment not	5 4	NI/A
	influenced by exposure status?	5.4	IN/A
	5.5. In diagnostic study, were test results blinded to patient history and other test	5.5	N/A
6	were <u>Intervention</u> /therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?	6	Yes
	6.1. In RCT or other intervention trial, were protocols described for all regimens	6.1	Yes
	studied?		
	6.2. In observational study, were interventions, study settings, and	6.2	N/A
	clinicians/provider described?	6.3	Yes
	6.3. Was the intensity and duration of the intervention or exposure factor sufficient		
	6.4 Was the amount of exposure and if relevant subject/patient compliance	6.4	Yes
	measured?	6.5	Yes
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.6	N/A
	6.6. Were extra or unplanned treatments described?		
	6.8. In diagnostic study, were details of test administration and replication	6.7	N/A
	sufficient?	6.8	N/A
7	Were outcomes clearly defined and the measurements valid and reliable?	7	Unclear
	7.1. Were primary and secondary endpoints described and relevant to the	7.1	Yes
	question ?	7.2	N/A
	7.3. Was the period of follow-up long enough for important outcome(s) to occur?	7.2	
	7.4. Were the observations and measurements based on standard, valid, and	1.3	IN/A
	reliable data collection instruments/tests/procedures?	7.4	Yes
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Unclear
	7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	Unclear
	7.7. Were the measurements conducted consistently across groups?	7.7	N/A
8	Was the statistical analysis appropriate for the study design and type of outcome	8	Unclear
	indicators?		
	8.1. Were statistical analyses adequately described the results reported	8.1	Unclear
	appropriately? 8.2 Were correct statistical tests used and assumptions of test not violated?	8.2	Yes
	one in the context statistical tests used and assumptions of test not volated;		

8.3.	Were statistics reported with levels of significance and/or confidence intervals?	8.3	Yes	
8.4.	Was "intent to treat" analysis of outcomes done (and as appropriate, was there	Q /	No	
	analysis of outcomes for those maximally exposed of a dose-response analysis)?	0.4	INU	
8.5.	Were adequate adjustments made for effects of confounding factors that	8.5	No	
	might have affected the outcomes (e.g., multivariate analyses)?	8.6	Ves	
8.6.	Was clinical significance as well as statistical significance reported?	0.0	105	
8.7.	If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A	
9. Are <u>co</u>	nclusions supported by results with biases and limitations taken into	9	Yes	
consic	consideration?		Ves	
9.1.	Is there a discussion of findings?	7.1	103	
9.2.	Are biases and study limitations identified and discussed?	9.2	Yes	
10. Is bias	due to study's <u>funding or sponsorship</u> unlikely?	10	Yes	
10.1	. Were sources of funding and investigators' affiliations described?	10.1	Yes	
10.2	. Was there no apparent conflict of interest?	10.2	Yes	

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

## PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

as found in JADA.       Initiality, criticity, trackers, a cross-sectional study. BMC Public Health, 17, 880.         doi:10.1186/s12889-017-4888-1       Cross-sectional study         Study design: Use algorithm – RCT, cohort, etc       Cross-sectional study         Study Class (A,B,C,D)       D         Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)       POSITIVE (+)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
as joint in order.       Weindow during duckers, across sectional study. Differ 1 able Frank, Fr, 660.         doi:10.1186/s12889-017-4888-1         Study design: Use algorithm – RCT, cohort, etc       Cross-sectional study         Study Class (A,B,C,D)       D         Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)       POSITIVE (+)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker
Study design: Use algorithm –       Cross-sectional study         RCT, cohort, etc       Cross-sectional study         Study Class (A,B,C,D)       D         Research Quality Rating       POSITIVE (+)         This rating tells if the research       design is good (+), bad (-) or         neutral (Ø)       This is determined by the quality       eventor         This is determined by the quality       eventor         criteria list. Delete the ratings       eventor         that do not apply (i.e. if positive,       eventor         delete minus/negative and       eventor         neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the       To explore users' experience of activity trackers, including the perceived         usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
Study (Lass) (, etc.)       Cross-sectional study         Study Class (A,B,C,D)       D         Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)       POSITIVE (+)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
Study Class (A,B,C,D)       D         Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)       POSITIVE (+)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker
Study Class (N, b, c, b)       D         Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)       POSITIVE (+)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker
Research Quality Rating This rating tells if the research design is good (+), bad (-) or neutral (Ø)FOSTIVE (+)This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).Purpose/Population Studied/Practice StudiedResearch purpose: What is the research question being investigated in the study?)Purpose/Population Studied/Practice StudiedInclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have for merty used an activity tracker
Inits rating tells if the research design is good (+), bad (-) or neutral (Ø)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).         Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)         Inclusion criteria: requirements for study eligibility         Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker smartphone app without an associated wearable activity
aesign is good (+), bud (+) or neutral (Ø)         This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).         Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)         Inclusion criteria: requirements for study eligibility         Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions
This is determined by the quality criteria list. Delete the ratings that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       Purpose/Population Studied/Practice Studied         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
Inits is determined by the quality         criteria list. Delete the ratings         that do not apply (i.e. if positive,         delete minus/negative and         neutral).         Purpose/Population Studied/Practice Studied         Research purpose: What is the         research question being         investigated in the study?)         Inclusion criteria: requirements         for study eligibility         Adults over 18 years of age, living in Australia, and either currently using or have         formerly used an activity tracker         Exclusion criteria (conditions
that do not apply (i.e. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
Induction upply (net. if positive, delete minus/negative and neutral).       Purpose/Population Studied/Practice Studied         Research purpose: What is the research question being investigated in the study?)       To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use         Inclusion criteria: requirements for study eligibility       Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker         Exclusion criteria (conditions       Use of an activity tracker smartphone app without an associated wearable activity
activity initial i
Purpose/Population Studied/Practice StudiedResearch purpose: What is the research question being investigated in the study?)To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to useInclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditionsUse of an activity tracker smartphone app without an associated wearable activity
Research purpose: What is the research question being investigated in the study?)To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to useInclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditionsUse of an activity tracker smartphone app without an associated wearable activity
Rescarch question being investigated in the study?)To explore users experience of activity dackets, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to useInclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditions)Use of an activity tracker smartphone app without an associated wearable activity
investigated in the study?)activity, diet, and sleep), ease of use, patterns of usages, and barriers to useInclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditions)Use of an activity tracker smartphone app without an associated wearable activity
Inclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditions)Use of an activity tracker smartphone app without an associated wearable activity
Inclusion criteria: requirements for study eligibilityAdults over 18 years of age, living in Australia, and either currently using or have formerly used an activity trackerExclusion criteria (conditions)Use of an activity tracker smartphone app without an associated wearable activity
Study englobility         Tormerry used an activity fracker           Exclusion criteria (conditions         Use of an activity tracker smartphone app without an associated wearable activity
<b>Exclusion criteria</b> (conditions Use of an activity tracker smartphone app without an associated wearable activity
<b>Exclusion criteria</b> (conditions ) Use of an activity tracker smartphone app without an associated wearable activity
that make individual ineligible) tracker use of a fitness watch which could not measure daily steps, and activity
trackers that cannot interact with a computer or smart phone
<b>Recruitment</b> Promoted using low-cost distribution methods on Facebook and Twitter primarily:
sharing the survey link with a variety of Facebook community groups in the field of
health and fitness (e.g. sporting clubs, cycling interest groups). Additionally, the
survey link was shared on the University of South Australia's Facebook and
Twitter feeds, and shared by individual members of the research team. An
incentive (a \$50 voucher random prize draw) was offered to encourage people to
share the social media posts. A second \$50 voucher random prize draw was
offered for people who completed the survey
Blinding used: some of the NA
persons involved are prevented
from knowing certain information
that might lead to conscious or
unconscious bias on their part,
invalidating the results
<b>Description of study protocol</b> A cross-sectional online survey was developed and administered to Australian
What happened in the study? adults who were current or former activity tracker users
Intervention: <i>Describe</i> A purpose-designed survey instrument was developed to address the research
<i>interventions, regimens, risk</i> objectives. The online survey was delivered via Survey Monkey. All participants
<i>factors, or procedures studied.</i> were asked for basic demographic characteristics, including sex, age, education
level, and relationship status. Participants were also asked whether they were
which brond of activity tracker of nad formerly used an activity tracker, and
which of and of activity tracker they currently or formerly used. In addition,
participants were asked now long they had worn their activity trackers, and if they were current users of an activity tracker, how long they intended to continue
were current users of an activity flacker, now long mey intended to continue wearing it into the future. Three items were included to assess how participants
used and shared the data derived from their activity trackers. Four items were
included to assess perceived behavior change related to use of the activity tracker

	Participants were asked to identify whether wearing their activity tracker motivated them in eight different domains, including: 'improve fitness', 'improve health', 'improve appearance', 'lose weight', 'monitor activities', 'share activities', 'compete with family or friends', and 'keep up with technology'. Up to three questions explored practical issues related to use of activity trackers. Finally, participants who had formerly worn an activity tracker were asked to select a reason why they had ceased to use it, from 10 options including reasons such as 'it broke', 'it was difficult to understand', or 'it wasn't helping with my goals', with an open-ended 'other' option included to capture additional reasons. Twenty-one items were included to assess current users' perceptions of the ease of use, usefulness, and accuracy of seven common features of activity trackers: active minutes, step counts, stair counts, sleep, heart rate, energy burned, and energy consumed. Responses were recorded on a 5-point Likert scale.
<b>Statistical analysis:</b> List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Categorical variables were analyzed using frequency of responses and percentages, and continuous variables were analyzed using medians, means, ranges, and standard deviations. Differences between former and current users were explored using independent samples t-tests, Mann-Whitney U tests, and chi square tests. Differences in use and experience on the basis of activity tracker manufacturer
<b>Timing of measurements:</b> when outcomes were measured; usually baseline and one or more later times	were also examined. Data collection took place in April to May 2016
<b>Dependent variables:</b> <i>outcomes</i> <i>that are measured or registered;</i> <i>variable whose change or</i> <i>different states the researcher</i> <i>wants to understand, explain, or</i> <i>predict</i>	Users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of use, and barriers to use
<b>Independent variables</b> (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Unclear survey questions, answers that do not apply
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	NA
<b>Initial n</b> (e.g. 731 (298 males, 433 females)) Record number that entered study – not the number screened.	305 participants
<b>Final n</b> (attrition) number of subjects that completed study	237 participants, 69 males and 168 females
Age usually mean or range	18 to 74 years
Ethnicity (if given)	Unclear
<b>Other relevant demographics:</b> <i>demographics describe the</i> <i>population (students, athletes, etc)</i>	52 (21.9%) participants completed high school, 42 (17.7%) participants completed technical and further education/certificate/diploma/apprenticeship, 95 (40.1%) participants earned an undergraduate degree, and 48 (20.3%) participants earned a postgraduate degree. 154 (65.0%) participants were in a relationship, 64 (27.0%)

	participants were single, and 19 (8.0%) participants did not specify their relationship status.				
Anthropometrics: e.g. were groups same or different on important physical measures (BMI, fitness level)	NA				
<b>Location:</b> <i>Where did the study</i> <i>take place? City or country</i>	Australia				
Summary of Results: Abstract results including quantitative data and statistics. Include statistical significance: P-values, confidence intervals (CI), relative risk (RR), odds ratios (OR), likelihood ratio, number needed to treat, power analysis if available.	Participants included 200 current and 37 former activity tracker users (total N = 237). The most commonly used brand of activity tracker was Fitbit (67.5%, n = 160), followed by Garmin (16.5%, n = 39), Apple (3.4%, n = 8), Jawbone (2.5%, n = 6), Samsung (1.7%, n = 4), Polar (1.3%, n = 3), and other (7.1%, n = 17). Participants typically used their activity trackers for sustained periods (5–7 months) and most intended to continue usage. The majority of current users either somewhat or strongly agreed that various features on their trackers were useful, including: steps (95%), active minutes (76%), sleep (66%), heart rate (63%), stairs climbed (58%), and energy burned (57%). Fewer agreed that the food intake feature was useful (36%). Participants reported they had improved their physical activity (51–81%) more commonly than they had their diet (14–40%) or sleep (11–24%), and slightly more participants reported to value the real time feedback (89%) compared to the long-term monitoring (78%). The majority of participants reported that they dia not use social features (65%) nor did they share their activity data on social media platforms (77%). A chi square test was conducted to determine whether participants who shared their activity tracker data via social media reported positive behavior change more frequently than participants who did not. The results suggested that they had had a positive experience using their activity tracker. A Mann-Whitney U test revealed this differed significantly, with current users more likely to report a positive experience than former users, U = 1746.50, z = -5.79, p = < 001, r = .38. Despite this, current users overall (U = 1648.5, z = -2.36, p = .02, r = .18). Former users were asked to identify why they no longer use their activity tracker. The main reasons given were that they felt they had learned everything they could from their tracker (30%), their tracker was broken (22%), and/or their tracker varied by brand. Only Fibit and Garmin were included in these comparisons, since ot				
Author conclusion: paraphrase	Findings suggest that in general, activity trackers are used for a substantial period				
that stated by study author in body of the report or abstract	of time and are viewed positively by users. Participants predominantly use their trackers to monitor and intervene on physical activity rather than other daily activities (e.g. sleep and diet) and were slightly more likely to value the trackers' real-time feedback more than long term monitoring capabilities. The majority of users perceived they had increased their physical activity as a result of using the activity tracker. Key barriers to continued use were device breakage or loss, and technical difficulties with the device or accompanying software. Findings support				
	activity trackers as appealing and useful tools for intervening on physical activity.				
Reviewer comments: Note	Strengths: large sample size, study specifically explored activity tracker users'				
-------------------------------------	------------------------------------------------------------------------------------	--	--	--	--
strengths and limitations of study;	perspectives, survey instrument was well-designed using a rigorous process,				
identify concerns that affect study	feedback from independent experts in the field, and underwent pilot testing				
validity and generalizability—	Limitations: relatively high dropout rate, predominantly female sample, study				
your comments should be	design increased the risk of recall bias, and difficulty knowing how generalizable				
italicized)	the results are				
	Funding source: The authors' have no funding to declare. The authors declare that				
	they have no competing interests.				

Table 3.2.a. Quality Criteria Checklist: Primary Research

DELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet) Maher, C., Ryan, J., Ambrosi, C., & Edney, S. (2017). Users' experiences of wearable activity		Y E S	N O	U N C	N A
trackers: a cross-sectional study. <i>BMC Public Health</i> , <i>17</i> , 880. doi:10.1186/s12889-017-4888-1				L E A R	
1. Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (Not Applicable for some anidomiclogical studies)	1			X	
<ul> <li>2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/ population group would care about?</li> </ul>	2	Х			
<ul> <li>3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?</li> </ul>	3		Х		
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "ves", the report is eligible for design	ation w	vith a	plu.	s (+)	) on
the Evidence Quality Worksheet, depending on answers to the following validity questions.			1	( )	
VALIDITY QUESTIONS					
1. Was the <u>research question</u> clearly stated?		Y E	1	N U D N	I N A
This is usually stated at end of the introduction and just before methods section.		s			1
		Х	K	E A R	2
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified? <i>This is often called the treatment and explained in the methods section.</i>	1.1	Х			
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Х	C		
These are sometimes called the endpoints; the results section reports the outcomes, but this information should be in the methods section, too					
1.3 Were the target population and setting specified?	1.3	X	C		
The target population is group for whom findings may be applicable; look for this in the introduction and in the methods section					
2. Was the <u>selection</u> of study subjects/patients free from bias?		Y	1		J N
		s		C	
				E	i i
		X		F	i l
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and without omitting criteria critical to the study?	2.1	Х			
The authors should give several points about the inclusion/exclusion criteria such as the age range of the subjects, disease condition (like hyperlipidemia) required for inclusion. Exclusion criteria should be listed, too, although some are					
understood. For example if the ages for inclusion are 18 to 70, the authors will probably not specifically note that children and people over age 70 were					
excluded. Most of the time, however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD).					
2.2 Were criteria applied equally to all study groups?	2.2	Х	ζ.		
2.3 Were health, demographics, and other characteristics of subjects described?	2.3	Х	Č.		
There is usually a Table 1 summarizing demographics and characteristics at baseline.					
Groups are <u>not</u> different if the P-Value is $> 0.05$ . If there has been a previous					
paper describing the study population, that paper may be referenced and you would need to go back to the aviating multipation to go that Table 1					
would need to go back to the original publication to see that Table 1.	2.4	v	,	+	—
<i>The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may</i>	2.4				
only say that the patients came from the same clinic from people who met the inclusion					
criteria.					

3. V	Vere <u>study groups comparable</u> ?		Y E	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		ŝ	Ŭ	C	
	Groups are <u>not</u> different if the P-Value is $> 0.05$ .				E	
			Х		A R	
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2	Х			
	demographics) similar across study groups at baseline? See Table I for this - there					
	should be no significant differences across study groups in an intervention study.					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3				Х
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
	prior study as a control group; that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control.					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4	Х			
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test was there an independent blind comparison with an appropriate	36				X
	reference standard (e.g. "gold standard")?	2.0				
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4	Was method of handling withdrawals described?		Y	N	U	N
	was method of handning wand avails described.		E S	0	N C	A
					L	
			x		A	
	4.1 Were follow up methods described and the same for all groups?	<i>A</i> 1			к	x
	4.1 Were the number characteristics of with drawels (i.e. dramouts lost to follow up	4.1	v			Λ
	4.2 was the number, characteristics of withdrawais (i.e. dropouts, lost to follow up,	4.2	Λ			
	attrition rate) and/or response rate (cross-sectional studies) described for each group?					
	(Follow up goal for a strong study is $80\%$ .)					
	This should be jound in the results section. If there is durition $\geq 20\%$ , it is important to					
	note that on the worksheet (as a note in the results section or in the reviewer		1			
	4.2 Ware all arrelled subjects/rections (in the arisinal seconds)	4.2		v		
	4.5 were an enioned subjects/patients (in the original sample) accounted for?	4.3	1	A		
	<i>This information is often presented in a figure with # recruited, # enrolled (Inis is Initial N). # non-aining at and of attack, point of (Court N). Sometime of the mercent of the line of the mercent of the second of the seco</i>		1			
	<i>N)</i> , <i>#</i> remaining at ena of study period (final N). Sometimes the reasons that subjects with draw on word drawing drawing in the formation in the formation of the text (county and its set).		1			
	wunarew or were aropped is given in the jigure or in the text (results section).	1.4	37			
	4.4 were reasons for windrawais similar across groups?	4.4	A			
	if there is a targe attrition from one group and not others, you would want to look for a		1			
	reason wny; the answer to this question would then be no.					

	4.5 If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4.5				Х
	The test under study should be compared to reference test all the time. An example of this might be using a DEXA machine to measure percent body fat only if a subject's					
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30%$ .					
5.	Was blinding used to prevent introduction of bias?		Y	N	U	N
0.			E S	0	C N	Α
					L E	
					A	Х
	5.1. In intermention study, more subjects, eliminican (unsetition and investigations blinded	5 1			~	v
	5.1 In Intervention study, were subjects, clinicians/practitioners and investigators blinded to tractment group as appropriate?	5.1				Λ
	The key term is as appropriate For example in the Lim et al 2008 study the investigators					
	studied the effect of MNT on linid levels in hypercholesterolemic nations. It was					
	an RCT, but obviously, the subjects and practitioners knew who was getting MNT					
	and who was not Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the dietitians and patients to know they were receiving MNT.					
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2			Х	
	using an objective test, such as a lab value, this criterion is assumed to be met.)					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3			Х	
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	Influenced by exposure status?					
	<i>Establish who is a case and who is a control at the beginning of the study.</i>	5.5				v
6	S.5 in diagnostic study, were test results blinded to patient instory and other test results?	5.5	Y	N	U	Λ N
0.	comparison(s) described in detail? Were intervening factors described?		E	0	N C	Α
	comparison(s) described in detail. Were intervening factors described.		5		L	
			x		A	
	6.1 In RCT or other intervention trial were protocols described for all regimens studied?	61			к	X
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2	x			
	described?	0.2	Λ			
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	63	x			
	produce a meaningful effect?	0.2				
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
	difference in lab values for cholesterol; however, 12 days would not be long					
	enough)					
	6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.4				Х
	How long did the treatment last? Did the patient follow directions?					
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5				Х
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)					
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					
	discussion section. It is likely there were no unplanned treatments, so a "no"					
	answer is not a problem overall.	67	v			
1	<i>i</i> , <i>i</i> was the information for 0.4, 0.5, 0.0 and 0.7 assessed the same way for all groups?	0./	Λ			
	6.8 In diagnostic study, were details of test administration and replication sufficient?	6.8				v
1	Usually answer n/a for diet study	0.0				Λ
	Sharry and the for all blady.	1	1	1	1	

7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E	N A
			Х		A R	
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability.</i>	7.5	Х			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6			X	
	7.7 Were the measurements conducted consistently across groups?	7.7	Х			
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E	N A
			X		R	
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	Х			
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	Х			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals ( <i>mean</i> $\pm$ <i>CI</i> )	8.3	X			
	8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)? Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i> . If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.	8.4				X
	8.5 Were adequate adjustments made for effects of confounding factors that might have	8.5		Х		
	affected the outcomes (e.g. multivariate analyses)? Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.					
	<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical</li> </ul>	8.6	X			

	significance is reported. Reducing cholesterol from 300 to 250 might be statistically significant, but clinically it is still abnormal.		T			
	<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li><i>Type II (β error is a false negative that happens when the investigators fail to reject the</i> <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				X
9.	Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S X	N O	U N C L E A P	N A
	9.1 Is there a discussion of findings? Answer yes or no.	9.1	X		ĸ	
	9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	X			
10.	<b>Is bias due to study's <u>funding or sponsorship</u> unlikely?</b> Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
	<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X			
	10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.	10.2	X			
SY	MBOL					
MI If m min	NUS/NEGATIVE (-) nost (six or more) of the answers to the above validity questions are "no," the report should a us (-) symbol on the Evidence Quality Worksheet. UTRAL (0)	be design	ated	with	h a	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

## PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Academy of Nutrition and Dietetics Evidence Analysis Library® Worksheet Template and Quality Criteria Checklist: Primary Research

Citation	tion Maher, C., Ryan, J., Ambrosi, C., & Edney, S. (2017). Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health, 17, 880. doi:10.1186/s12889-017-4888-1						
Study Design	Cross-sectional study						
Class	D						
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\bigcirc$ (Neutral)						
Research Purpose	To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of usages, and barriers to use						
Inclusion Criteria	Adults over 18 years of age, living in Australia, and either currently using or have formerly used an activity tracker						
Exclusion Criteria	Use of an activity tracker smartphone app without an associated wearable activity tracker, use of a fitness watch which could not measure daily steps, and activity trackers that cannot interact with a computer or smart phone						
Description of Study Protocol	Recruitment: Promoted using low-cost distribution methods on Facebook and Twitter, primarily; sharing the survey link with a variety of Facebook community groups in the field of health and fitness (e.g. sporting clubs, cycling interest groups). Additionally, the survey link was shared on the University of South Australia's Facebook and Twitter feeds, and shared by individual members of the research team. An incentive (a \$50 voucher random prize draw) was offered to encourage people to share the social media posts. A second \$50 voucher random prize draw was offered for people who completed the survey Design: A cross-sectional online survey was developed and administered to Australian adults who were current or former activity tracker users Blinding used (if applicable): NA Intervention (if applicable): A purpose-designed survey instrument was developed to address the research objectives. The online survey was delivered via Survey Monkey. All participants were asked for basic demographic characteristics, including sex, age, education level, and relationship status. Participants were also asked whether they were						

currently using an activity tracker or had formerly used an activity tracker, and which brand of activity tracker they currently or formerly used. In addition, participants were asked how long they had worn their activity trackers, and if they were current users of an activity tracker, how long they intended to continue wearing it into the future. Three items were included to assess how participants used and shared the data derived from their activity trackers. Four items were included to assess perceived behavior change related to use of the activity tracker. Participants were asked to identify whether wearing their activity tracker motivated them in eight different domains, including: 'improve fitness', 'improve health', 'improve appearance', 'lose weight', 'monitor activities', 'share activities', 'compete with family or friends', and 'keep up with technology'. Up to three questions explored practical issues related to use of activity trackers. Finally, participants who had formerly worn an activity tracker were asked to select a reason why they had ceased to use it, from 10 options including reasons such as 'it broke', 'it was difficult to understand', or 'it wasn't helping with my goals', with an open-ended 'other' option included to capture additional reasons. Twenty-one items were included to assess current users' perceptions of the ease of use, usefulness, and accuracy of seven common features of activity trackers: active minutes, step counts, stair counts, sleep, heart rate, energy burned, and energy consumed. Responses were recorded on a 5-point Likert scale.

Statistical Analysis: Categorical variables were analyzed using frequency of responses and percentages, and continuous variables were analyzed using medians, means, ranges, and standard deviations. Differences between former and current users were explored using independent samples t-tests, Mann-Whitney U tests, and chi square tests. Differences in use and experience on the basis of activity tracker manufacturer were also examined.

Data Collection Summary	Timing of Measurements: Data collection took place in April to May 2016 Dependent Variables: Users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of use, and barriers to use Independent Variables: Unclear survey questions, answers that do not apply Control Variables: NA
Description of Actual Data Sample	Initial: 305 (Males Females) Attrition (final N): 237 (69 males and 168 females) Age: 18 to 74 years Ethnicity: Unclear Other relevant demographics: 52 (21.9%) participants completed high school, 42 (17.7%) participants completed technical and further education/certificate/diploma/apprenticeship, 95 (40.1%) participants earned an undergraduate degree, and 48 (20.3%) participants earned a postgraduate degree. 154 (65.0%) participants were in a relationship, 64 (27.0%) participants were single, and 19 (8.0%) participants did not specify their relationship status. Anthropometrics: NA Location: Australia
Summary of Results	Key Findings: Participants included 200 current and 37 former activity tracker users (total N = 237). The most commonly used brand of activity tracker was Fitbit (67.5%, n = 160), followed by Garmin (16.5%, n = 39), Apple (3.4%, n = 8), Jawbone (2.5%, n = 6), Samsung (1.7%, n = 4), Polar (1.3%, n = 3), and other (7.1%, n = 17). Participants typically used their activity trackers for sustained periods (5–7 months) and most intended to continue usage. The majority of current users either somewhat or strongly agreed that various features on their trackers were useful, including: steps (95%), active minutes (76%), sleep (66%), heart

rate (63%), stairs climbed (58%), and energy burned (57%). Fewer agreed that the food intake feature was useful (36%). Participants reported they had improved their physical activity (51-81%) more commonly than they had their diet (14-40%) or sleep (11-24%), and slightly more participants reported to value the real time feedback (89%) compared to the long-term monitoring (78%). The majority of participants reported that they did not use social features (65%) nor did they share their activity data on social media platforms (77%). A chi square test was conducted to determine whether participants who shared their activity tracker data via social media reported positive behavior change more frequently than participants who did not. The results suggested that sharing data via social media was not associated with behavior change, X2 (1) = 1.07, p = .30. Overall, 94% of current users and 65% of former users agreed that they had had a positive experience using their activity tracker. A Mann-Whitney U test revealed this differed significantly, with current users more likely to report a positive experience than former users, U = 1746.50, z = -5.79, p = <.001, r = .38. Despite this, current users reported technical issues or other complaints relating to their activity trackers, most commonly relating to the tracker not suiting their outfit (19%), low battery life (19%), difficulties with the support software (17%), or perceived inaccuracy of data collected (17%). Former users reported more issues than current users overall (U = 1648.5, z = -2.36, p = .02, r = .18). Former users were asked to identify why they no longer use their activity tracker. The main reasons given were that they felt they had learned everything they could from their tracker (30%), their tracker was broken (22%), and/or their tracker was not helping them achieve their goals (14%). Finally, analyses were performed to determine whether users' experiences and perceptions relating to activity tracker varied by brand. Only Fitbit and Garmin were included in these comparisons, since other brands had insufficient sample sizes. The perceived usefulness and accuracy of activity data did not vary between

	brands. However, some aspects of ease of use did vary. Fitbit users rated					
	the stair climbing, heart rate, and dietary intake features as being					
significantly easier to use than Garmin users did ( $p = 0.01-0.02$ ).						
	Other Findings:					
	Findings suggest that in general, activity trackers are used for a substantial					
	period of time and are viewed positively by users. Participants					
	predominantly use their trackers to monitor and intervene on physical					
	activity rather than other daily activities (e.g. sleep and diet) and were					
	slightly more likely to value the trackers' real-time feedback more than					
Author	long term monitoring capabilities. The majority of users perceived they					
Conclusion	had increased their physical activity as a result of using the activity					
	tracker. Key barriers to continued use were device breakage or loss, and					
	technical difficulties with the device or accompanying software. Findings					
	support activity trackers as appealing and useful tools for intervening on					
	physical activity.					
	Strengths: large sample size, study specifically explored activity tracker					
	users' perspectives, survey instrument was well-designed using a rigorous					
	process, feedback from independent experts in the field, and underwent					
Reviewer	pilot testing					
Comments	Limitations: relatively high dropout rate, predominantly female sample,					
	study design increased the risk of recall bias, and difficulty knowing how					
	generalizable the results are					
	The authors' have no funding to declare. The authors declare that they					
Funding Source	have no competing interests.					

# Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
I	<b>Positive</b> – Indicates that the report has clearly addressed issues of
÷	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
0	exceptionally week

# Select a rating from the drop-down menu $\checkmark$

Relevance Questions								
<ol> <li>Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)</li> </ol>	1	Unclear						
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes						
3. Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No						
4. Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes						
If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.								
Validity Questions								
1. Was the <u>research question</u> clearly stated?	1	Yes						
1.1. Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes						
1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes						
1.3. Were the target population and setting specified?	1.3	Yes						
2. Was the <u>selection</u> of study subjects/patients free from bias?	2	Yes						
2.1. Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes						
without omitting criteria critical to the study?	2.2	Yes						
2.3. Were health, demographics, and other characteristics of subjects described?	2.3	Yes						
2.4. Were the subjects/patients a representative sample of the relevant population?	2.4	Yes						
<ul> <li>Were study groups comparable?</li> <li>3.1. Was the method of assigning subjects/patients to groups described and unbiased? (Method of randomization identified if PCT)</li> </ul>	3	Yes						
<ul> <li>3.2. Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?</li> </ul>	3.1	N/A						
<ul> <li>3.3. Were concurrent controls used? (Concurrent preferred over historical controls.)</li> </ul>	3.2	Yes						
3.4. If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.3	N/A						
<ul> <li>appropriate adjustments in statistical analysis?</li> <li>3.5. If case control study, were potential confounding factors comparable for cases and controls? (If case series or trial with subjects serving as own control, this</li> </ul>	3.4	Yes						
criterion is not applicable. Criterion may not be applicable in some cross- sectional studies.)	3.5	N/A						
3.6. If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	N/A						
4. Was method of handling <u>withdrawals</u> described?	4	Yes						
4.1. Were follow up methods described and the same for all groups?	4.1	N/A						
	4.2	Yes						

		4.2. Was the number,	characteristics of withdrawals (i.e., dropouts	s, lost to follow	4.3	No
		up, attrition rate)	and/or response rate (cross-sectional studie www.goal.for.a.strong.study.is.80%)	es) described for	4.4	Yes
		4.3. Were all enrolled	subjects/patients (in the original sample) ac	counted for?		
		4.4. Were reasons for	withdrawals similar across groups		4.5	
		4.5. If diagnostic test,	was decision to perform reference test not o	dependent on	4.5	N/A
		results of test und	er study?			
	5. V	as <u>blinding</u> used to pre	vent introduction of bias?	adiovectigators	5	N/A
		blinded to treatm	ent group, as appropriate?	and investigators		
		5.2. Were data collect	ors blinded for outcomes assessment? (If ou	tcome is	5.1	N/A
		measured using a	n objective test, such as a lab value, this crit	erion is assumed	52	Unclear
		to be met.)		<b>6</b>	5.2	Olicical
		5.3. In cohort study or risk factors blinde	cross-sectional study, were measurements d?	of outcomes and	5.3	Unclear
		5.4. In case control stu	dy, was case definition explicit and case asc	ertainment not	54	N/A
		Influenced by exp	<pre>&gt;Sure status? / wore test results blinded to patient bistor</pre>	wand other test		11/11
		results?	, were test results binded to patient history		5.5	N/A
	6. V	ere <u>intervention</u> /thera mparison(s) described	peutic regimens/exposure factor or proced in detail? Were intervening factors describ	ure and any ed?	6	Yes
		6.1. In RCT or other in	ervention trial, were protocols described fo	r all regimens	6.1	N/A
		studied?		-	( )	X7
		6.2. In observational s	udy, were interventions, study settings, and	ł	6.2	Yes
		Clinicians/provide	Oescribed?	e factor sufficient	6.3	Yes
		to produce a mea	ningful effect?		6.4	N/A
		6.4. Was the amount o	of exposure and, if relevant, subject/patient	compliance	6.5	NI/A
		measured? 6.5 Were co-interven	ions (e.g. ancillary treatments other therai	nies) described?	0.5	IN/A
		6.6. Were extra or unp	lanned treatments described?		6.6	N/A
		6.7. Was the informat	on for 6.4, 6.5, and 6.6 assessed the same w	vay for all groups?	6.7	Yes
		6.8. In diagnostic stud sufficient?	In diagnostic study, were details of test administration and replication	6.8	N/A	
-	7 14	ere outcomes clearly d	efined and the measurements valid and rel	iahle?	7	Vaa
	<i>.</i>	7.1. Were primary and	secondary endpoints described and relevant	nt to the	/	res
		question?			7.1	Yes
		7.2. Were nutrition me	asures appropriate to question and outcom	nes of concern?	7.2	N/A
		7.3. Was the period of	follow-up long enough for important outcome tions and manyuraments based on standard	me(s) to occur?	7.3	N/A
		reliable data colle	ction instruments/tests/procedures?	, valiu, anu	7.4	Yes
		7.5. Was the measure	nent of effect at an appropriate level of pre-	cision?	7.5	Yes
		7.6. Were other factor	s accounted for (measured) that could affec	t outcomes?	7.6	Unclear
		7.7. Were the measure	ements conducted consistently across group	IS?	7.7	Yes
	8. V	as the <u>statistical analys</u>	is appropriate for the study design and typ	e of outcome	8	Yes
	ir	dicators?	adverse adaquately described the results rea	ortod	0 1	Vas
		appropriately?	aryses adequately described the results rep	orteu	0.1	1 05
		8.2. Were correct stat	stical tests used and assumptions of test not	t violated?	8.2	Yes
		8.3. Were statistics re	ported with levels of significance and/or con	fidence intervals?	8.3	Yes

-				
	8.4. Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response		8.4	N/A
		analysis)?	8.5	No
	8.5.	Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g., multivariate analyses)?	8.6	Yes
	8.6. 8.7.	Was clinical significance as well as statistical significance reported? If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A
9.	Are <u>co</u>	nclusions supported by results with biases and limitations taken into	9	Yes
	consic	leration?	91	Ves
	9.1.	Is there a discussion of findings?	7.1	105
	9.2.	Are biases and study limitations identified and discussed?	9.2	Yes
10.	Is bias	due to study's <u>funding or sponsorship</u> unlikely?	10	Yes
	10.1	. Were sources of funding and investigators' affiliations described?	10.1	Yes
	10.2	. Was there no apparent conflict of interest?	10.2	Yes

#### MINUS/NEGATIVE (-)

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

#### NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Naslund, J., Aschbrenner, K., Scherer, E., McHugo, G., Marsch, L., & Bartels, S. (2016). Wearable Devices and Mobile Technologies for Supporting Behavioral Weight Loss Among People with Serious Mental Illness. <i>Psychiatry Research</i> , 244, 12014 (2016).
	139–144. doi:10.1016/j.psychres.2016.06.056
Study design: Use algorithm – RCT, cohort, etc	Non-randomized crossover trial
Study Class (A,B,C,D)	С
Research Quality Rating	POSITIVE (+)
This rating tells if the research	
neutral ( $\emptyset$ )	
This is determined by the quality	
criteria list. Delete the ratings	
that do not apply (i.e. if positive,	
delete minus/negative and	
neutral).	
	Purpose/Population Studied/Practice Studied
Research purpose: What is the	To examine whether average daily step count measured using Fitbit Zip wearable
research question being	devices was associated with weight loss and improved fitness among individuals
investigated in the study?)	with serious mental liness enrolled in a 6-month litestyle program
Inclusion criteria: requirements	Participants were age 21 or older; had serious mental illness defined by a diagnosis
for study eligibility	of schizophrenia, schizoaffective disorder, major depressive disorder, or bipolar
	disorder, spoke English, were on stable pharmacological treatment defined as
	obesity defined as body mass index (BMI) $> 30$
Exclusion criteria (conditions	Participants with any medical contraindication to weight loss: were pregnant or
that make individual ineligible)	planning to become pregnant within the next six months: or had a current diagnosis
	of an active alcohol-use or substance-use disorder
Recruitment	All participants were receiving services through community mental health settings
Blinding used: some of the	NA
persons involved are prevented	
from knowing certain information	
that might lead to conscious or	
unconscious bias on their part,	
invalidating the results	Destricted a schier share is successed in a size descent in diseases and
What happened in the study?	bipolar disorder, and wore Fitbit Zips most of the days they were enrolled in the 6-
	month lifestyle program
Intervention: Describe	Participants were enrolled in a 6-month group behavioral weight loss program
interventions, regimens, risk	targeting fitness and healthy eating through an urban community mental health
factors, or procedures studied.	center. The program was modeled after the evidence-based Diabetes Prevention
	Program, and included weekly group sessions led by lifestyle coaches. Prior to
	starting the program, participants received medical clearance from a primary care
	provider. Participants were given Fitbit Zip wearable devices and smartphones to
	use for the o-month study duration. Participants attended two brief 30-minute
	wearable device and synching the Fithit Zin with the associated smartphone
	application Technical support for using the Fithit Zip or associated application
	was provided to participants on an as needed basis by a member of the research
	team over the study duration. The goal-setting component of the program was
	personalized to meet participants' physical abilities. In general, participants shared
	an activity goal of reaching 150 minutes of exercise each week. As part of the

	program, participants received 3-5 text messages from research staff each week as reminders to attend optional exercise classes, to be more active as part of their daily routines, to provide encouragement, and to support participants in reaching the program's weekly physical activity goal. Daily step count data for the 6-month study duration was exported from participants' personal Fitbit accounts into an Excel spreadsheet. Fitness was measured using the 6-Minute Walk Test (6-MWT), which measures the distance in feet that an individual can walk in six minutes. Change in fitness was calculated as the change in feet on the 6-MWT from baseline to 6-months. Weight was measured and reported as the change in body weight from baseline to 6-months.
Statistical analysis: List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Linear regression models were used to evaluate the association between average daily step count over the 6-month study duration and the pre-post changes in participants' weight and fitness. Penalized functional regression models were used to evaluate the time-varying association between daily step count collected and the pre-post changes in participants' weight and fitness. A p-value of 0.05 was considered statistically significant.
<b>Timing of measurements:</b> when outcomes were measured; usually baseline and one or more later times	Daily step count data was collected continuously throughout the 6-month study duration. Body weight and fitness were measured at baseline and after 6-months.
<b>Dependent variables:</b> <i>outcomes</i> <i>that are measured or registered;</i> <i>variable whose change or</i> <i>different states the researcher</i> <i>wants to understand, explain, or</i> <i>predict</i>	Changes in step count data according to the Fitbit Zip, changes in body weight and fitness after the 6-month intervention
<b>Independent variables</b> (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Goal-setting component of the weight loss program, weekly group sessions led by lifestyle coaches targeting fitness and healthy eating, and several weekly text messages from research staff
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	Body weight and fitness at baseline
Initial n (e.g. 731 (298 males, 433 females)) Record number that entered study – not the number screened.	43 participants
<b>Final n</b> (attrition) number of subjects that completed study	34 participants (13 males and 21 females)
Age usually mean or range	Mean age: 50.2 years, SD = 11.0
Ethnicity (if given)	Non-Hispanic white
Other relevant demographics: demographics describe the population (students, athletes, etc)	Clinical characteristics: eight participants (23.5%) had a schizophrenia spectrum disorder, 17 participants (50.0%) had major depressive disorder, and nine participants (26.5%) had bipolar disorder. Education: two participants (5.9%) had less than a high school education, 11 participants (32.4%) attended high school, 13 participants (38.2%) attended some college, and eight participants (23.5%) earned a college degree. Living situation: 27 participants (79.4%) lived independently, six participants (17.7%) lived with family, and one participant (2.9%) lived in supported housing/assisted living. Marital status: 11 participants (32.4%) were

Anthronomotriou o o unro	never married, three participants (8.8%) were currently married, and 20 participants (58.8%) were previously married. Employment/Insurance coverage: seven participants (20.6%) were currently employed (part or full-time), eight participants (23.5%) were enrolled in Medicaid only, eight participants (23.5%) were enrolled in Medicaid only, eight participants (23.5%) were enrolled in Medicaid only, and 18 participants (52.9%) were dual eligible (enrolled in Medicaid and Medicare). Six participants (17.7%) were current smokers.
Anthropometrics: e.g. were groups same or different on important physical measures (BMI, fitness level)	Mean Weight: 231.9 pounds, $SD = 46.7$ Mean BMI: 38.5 kg/m2, $SD = 9.3$ Fitness: 1303.8 feet in 6-MWT, $SD = 323.2$
Location: Where did the study take place? City or country	Urban community mental health center in southern New Hampshire
Summary of Results: Abstract results including quantitative data and statistics. Include statistical significance: P-values, confidence intervals (CI), relative risk (RR), odds ratios (OR), likelihood ratio, number needed to treat, power analysis if available.	Due to rolling enrollment at the start of the study, participants had their Fitbit Zips to use for an average of 181.7 days (SD = 34.7), with a median of 181.5 days (interquartile range = 169 to 196). Participants wore their Fitbit Zips for a mean of 86.2% (SD = 18.4%) of the days that they had the Fitbit Zip to use (median = 94.0% of the days, interquartile range = 82.0% to 97.0%). Participants achieved an average of 4453.5 (SD = 2707.4) steps each day, with average daily step counts ranging from 1037.6 (SD = 767.9) steps to 11,366.3 (SD = 3416.9) steps. In total, 21 participants (61.8%) achieved 10,000 steps or more on at least one day during the study. These 21 participants achieved 10,000 daily steps or more for a mean of 16.1% (SD = 21.0%) of the days that they had the Fitbit Zip to use (median = 25.7% of the days, interquartile range = 0.9% to 25.7%). Three participants (8.8%) achieved 10,000 daily steps or more on at least they were enrolled in the study. There was a significant association between participants' average daily step count and weight loss. For every 1000 step increase in participants' aliy average step count, they experienced a decrease in weight of 1.78 pounds (F = 5.07; df = 1, 32; p = 0.0314). The relationship between average daily step count increased by 1000 steps, it corresponded to an increase of 18.79 feet on the 6-Minute Walk Test (F = 1.92; df = 1, 31; p = 0.176). In the penalized functional regression models, the time-varying relationship between daily step count and weight loss (permutation test statistic = 0.180; p = 0.264) and improved fitness (permutation test statistic = 0.076; p = 0.574) were not significant. This suggests that there was no specific period of time for which an increase in steps was significantly associated with either weight loss or improved fitness.
	Author's Conclusions
Author conclusion: paraphrase that stated by study author in body of the report or abstract Reviewer comments: Note	At 6-months, higher average daily step count was associated with greater weight loss, but not improved fitness. These findings suggest that wearable devices and their associated smartphone applications may serve as valuable tools for supporting community-based weight loss efforts for people with serious mental illness. Importantly, it was observed that a higher average daily step count over the 6- month program duration was associated with greater weight loss. Therefore, it appears that providing participants with serious mental illness the recommendation to collect more steps each day and maintain a high average daily step count throughout participation in a lifestyle intervention may contribute to greater weight loss. These are preliminary findings and should be interpreted with caution, but they offer promise regarding the potential benefits of using wearable devices to support lifestyle interventions delivered through community mental health settings. Strengths: <i>long study duration</i> , highly engaged participants. Fithit Zins were used
strengths and limitations of study; identify concerns that affect study validity and generalizability— your comments should be italicized)	to support self-monitoring, goal-setting, and tracking progress over time Limitations: <i>small sample size, predominantly female sample, lacking racial or</i> <i>ethnic diversity, relatively high dropout rate, blinding was not utilized,</i> analyses were based on the participants who completed the 6-month intervention, because Fitbit Zips were integrated as part of the behavioral weight loss program, the results cannot separate the impact of group education and support from use of the wearable

device, and all participants were receiving services through community mental
health settings, thus findings are likely not representative of individuals with
serious mental illness not currently receiving services
Funding source: This study was supported by the National Institute of Mental
Health (R01 MH089811-01) and the United States Centers for Disease Control and
Prevention Health Promotion and Disease Prevention Research Center
(Cooperative Agreement Number U48DP005018). The funders had no role in
study design, data collection and analysis, decision to publish, or preparation of the
manuscript. The authors report no conflicting interests.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)		Y E	N O	U N	N A
Naslund, J., Aschbrenner, K., Scherer, E., McHugo, G., Marsch, L., & Bartels, S. (2016).				L	
Decente vietes and Mobile Technologies for Supporting Benavioral weight Loss Among				E A	
doi:10.1016/j.mouchros.2016.06.056				R	
doi:10.1010/J.psychres.2010.00.030	1	v			
1. Would implementing the studied intervention or procedure (if found successful) result in	1	А			
anidomial agiaal studioa)					
2 Did the outborg study on outcome (dependent variable) or tonic that the nationts/slights/	2	v			
2. Did the authors study an outcome (dependent variable) of topic that the patients/chemis/	2	Λ			
3 Is the focus of the intervention or procedure (independent variable) or topic of study a	3	X			
common issue of concern to dietetics practice?	5				
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "ves", the report is eligible for design	ation w	ith a	nlus	(+)	on
the Evidence Quality Worksheet, depending on answers to the following validity questions.			p	( )	
VALIDITY QUESTIONS					
1. Was the research question clearly stated?		Y	Ň		N A
This is usually stated at end of the introduction and just before methods section.		S		C	A
				L E	
		X		AR	
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified?	11	v			-
This is often called the treatment and explained in the methods section	1.1				
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?		x			+
These are sometimes called the and points: the results section reports the outcomes but					
this information should be in the methods section too					
1.3 Were the target population and setting specified?	13	X			
The target population is group for whom findings may be applicable: look for this in the	1.5	1			
introduction and in the methods section					
2. Was the selection of study subjects/patients free from bias?		Y	N	U	N
		S		C	A
				L E	
		X		AR	
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1	X	:		1
diagnostic or prognosis criteria), and with sufficient detail and without omitting					
criteria critical to the study?					
The authors should give several points about the inclusion/exclusion criteria such as the					
age range of the subjects, disease condition (like hyperlipidemia) required for					
inclusion. Exclusion criteria should be listed, too, although some are					
understood. For example if the ages for inclusion are 18 to 70, the authors will					
probably not specifically note that children and people over age 70 were					
excluded. Most of the time, however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD).					
2.2 Were criteria applied equally to all study groups?	2.2				Х
2.3 Were health, demographics, and other characteristics of subjects described?	2.3	Х			
There is usually a Table 1 summarizing demographics and characteristics at baseline.					
Groups are <u>not</u> different if the P-Value is $> 0.05$ . If there has been a previous					1
paper describing the study population, that paper may be referenced and you					
would need to go back to the original publication to see that Table 1.			+	+	_
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4				
The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may					1
criteria.					

3. Were study groups comparable?		Y	N O	U N	N A
There is usually a Table 1 summarizing demographics and characteristics at baseline.		s	Ŭ	Ĉ	
<i>Groups are</i> <u>not</u> different if the <i>P</i> -Value is $> 0.05$ .				E	
		Х		A R	
3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
(Method of randomization identified if RCT)					
In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
randomized block design; or assigned by computer-generated random numbers).					
Look for instances that show bias; for example I once read a study where patients					
were randomized to receive liquid energy supplements; however, if someone					
disliked their supplement, they were allowed to change groups – this is not unbiased.					
3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2	Х			
demographics) similar across study groups at baseline? See Table I for this - there					
should be no significant differences across study groups in an intervention study.					
3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Х			
Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
prior study as a control group: that is an example of a historical control. That is not					
as strong a research design as use of concurrent control group A crossover study					
where the subject acts as his/her own control is use of concurrent control.					
3.4 If cohort study or cross-sectional study were groups comparable on important	34				X
confounding factors and/or were preexisting differences accounted for by using	5.1				11
appropriate adjustments in statistical analysis?					
The groups in a cohort or cross-sectional study should not be different from each other:					
if they are a strong study will utilize statistical techniques such as multivariate					
analyses to remove the variance due to the group differences. Look for this					
information in the statistics and results sections					
3.5. If case control study, were notential confounding factors comparable for cases and	3.5				x
controls? If case series or trial with subjects serving as own control, this criterion is	5.5				Λ
not applicable. Criterion may not be applicable in some cross-sectional studies					
Subjects are generally matched for age gender, etc. Look for this in the statistical					
description and results sections					
2.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6				v
reference standard (e.g. "gold standard")?	5.0				Λ
Example: comparing body fat analysis method with underwater weighing (gold					
standard) In studies trying to determine the best equation (like Mifflin St. Jear or					
Harris-Renedict) to predict energy needs, a gold standard measure of REE (Indirect					
Calorimetry) is used					
4 Was method of handling withdrewals described?		Y	N	U	N
4. Was method of handling <u>withdrawais</u> described:		E	0	N	А
		5		L	
		v		E A	
4.1 Ware follow up methods described and the same for all groups?	<u> </u>	Λ		R	v
4.1 were follow up methods described and the same for all groups?	4.1	37			Λ
4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up,	4.2	Х			
attrition rate) and/or response rate (cross-sectional studies) described for each group?					
(Follow up goal for a strong study is 80 %.)					
This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
note that on the worksheet (as a note in the results section or in the reviewer					
<i>comments at the very bottom)</i>	4.2		37		
4.5 were all enrolled subjects/patients (in the original sample) accounted for?	4.3				
<i>This information is often presented in a figure with # recruited, # enrolled (this is initial</i>					
N, # remaining at end of study period (final N). Sometimes the reasons that subjects					
withdrew or were dropped is given in the figure or in the text (results section).					
4.4 Were reasons for withdrawals similar across groups?	4.4	X			
If there is a large attrition from one group and not others, you would want to look for a					
reason why; the answer to this question would then be no.					

	<ul> <li>4.5 If diagnostic test, was decision to perform reference test not dependent on results of test under study?</li> <li>The test under study should be compared to reference test all the time. An example of this might be using a DEXA machine to measure percent body fat only if a subject's</li> </ul>	4.5				X
	BMI was > 35 but bioimpedance analyzer indicated body fat $< 30\%$ .					
5.	Was blinding used to prevent introduction of bias?		Y	N	UN	N
			S	0	C L	А
					E	ł
				x	A R	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded to treatment group, as appropriate?	5.1		X		
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was an RCT, but obviously, the subjects and practitioners knew who was getting MNT and who was not. Therefore, you would not answer question 5.1 NO. It was appropriate for the dietitians and patients to know they were receiving MNT.					
	<ul> <li>5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured using an objective test, such as a lab value, this criterion is assumed to be met.)</li> <li>Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet study is to have separate people analyzing the data (not the same ones who were collecting the data)</li> </ul>	5.2		X		
	<ul><li>5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk factors blinded?</li></ul>	5.3				Χ
	Answer ves, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were					
	collecting the data).					
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	influenced by exposure status?					
	Establish who is a case and who is a control at the beginning of the study.					
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5				X
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any		Y E	N O	U N	N A
	comparison(s) described in detail? Were intervening factors described?		s		C L	
			v		E A	
	(1 In DOT on other intervention trial more motocals described for all regimens studied)	6.1			R	
	6.1 In RC1 of other intervention that, were protocols described for an regimens studied?	0.1	Λ			V
	6.2 In observational study, were interventions, study settings, and clinicians/provider described?	6.2				X
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	Х			
	produce a meaningful effect?					
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					
	difference in lab values for cholesterol; however, 12 days would not be long					
	enougn)	6.4	v			
	How long did the treatment last? Did the patient follow directions?	0.4	Λ			
	6.5 Were co-interventions (e.g. ancillary treatments other therapies) described?	6.5	x			
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)					
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					
	discussion section. It is likely there were no unplanned treatments, so a "no"					
	answer is not a problem overall.	<				
1	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7				Х
	For a study to be valid and unblased, it is important that this be yes.	6.8				v
	Usually answer n/a for diet study.	0.0				Λ

7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E	N A
			Х		A R	
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability.</i>	7.5	Х			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A	N A
		0.1	A		R	
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	X			
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals ( <i>mean</i> $\pm$ <i>CI</i> )	8.3	Х			
	8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)? Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i> . If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.	8.4		X		
	8.5 Were adequate adjustments made for effects of confounding factors that might have	8.5		X		
	affected the outcomes (e.g. multivariate analyses)? Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.					
	<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical</li> </ul>	8.6	X			

_						
	significance is reported. Reducing cholesterol from 300 to 250 might be statistically significant, but clinically it is still abnormal.					
	<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li>Type II (β error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				X
9.	Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S X	N O	U N C L E A	N A
	9.1 Is there a discussion of findings? Answer yes or no.	9.1	X		ĸ	
	9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	X			
10.	<b>Is bias due to study's <u>funding or sponsorship</u> unlikely?</b> Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
	<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X			
	10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.	10.2	X			
SY	MBOL					
MI If n min	NUS/NEGATIVE (-) nost (six or more) of the answers to the above validity questions are "no," the report should a nus (-) symbol on the Evidence Quality Worksheet. UTRAL (0)	be design	nated	with	n a	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

## PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Citation	Naslund, J., Aschbrenner, K., Scherer, E., McHugo, G., Marsch, L., & Bartels, S. (2016). Wearable Devices and Mobile Technologies for Supporting Behavioral Weight Loss Among People with Serious Mental Illness. Psychiatry Research, 244, 139–144. doi:10.1016/j.psychres.2016.06.056
Study Design	Non-randomized crossover trial
Class	С
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\otimes$ (Neutral)
Research Purpose	To examine whether average daily step count measured using Fitbit Zip wearable devices was associated with weight loss and improved fitness among individuals with serious mental illness enrolled in a 6-month lifestyle program
Inclusion Criteria	Participants were age 21 or older; had serious mental illness defined by a diagnosis of schizophrenia, schizoaffective disorder, major depressive disorder, or bipolar disorder; spoke English; were on stable pharmacological treatment defined as receiving the same psychiatric medications over the prior two months; and had obesity defined as body mass index (BMI) $\geq$ 30
Exclusion Criteria	Participants with any medical contraindication to weight loss; were pregnant or planning to become pregnant within the next six months; or had a current diagnosis of an active alcohol-use or substance-use disorder
Description of Study Protocol	neerdinnent. An participants were receiving services through community mental health settings Design: Participants had a schizophrenia spectrum disorder, major depressive disorder, or bipolar disorder, and wore Fitbit Zips most of the days they were enrolled in the 6-month lifestyle program Blinding used (if applicable): NA Intervention (if applicable): Participants were enrolled in a 6-month group behavioral weight loss program targeting fitness and healthy eating through an urban community mental health center. The program was modeled after the evidence-based Diabetes Prevention Program, and included weekly group sessions led by lifestyle coaches. Prior to starting the program, participants received medical clearance from a primary care provider. Participants were given Fitbit Zip wearable devices and

	smartphones to use for the 6-month study duration. Participants attended
	two brief 30-minute training sessions with a member of the research staff
	for instruction in using the wearable device and synching the Fitbit Zip
	with the associated smartphone application. Technical support for using
	the Fitbit Zip or associated application was provided to participants on an
	as needed basis by a member of the research team over the study duration.
	The goal-setting component of the program was personalized to meet
	participants' physical abilities. In general, participants shared an activity
	goal of reaching 150 minutes of exercise each week. As part of the
	program, participants received 3-5 text messages from research staff each
	week as reminders to attend optional exercise classes, to be more active as
	part of their daily routines, to provide encouragement, and to support
	participants in reaching the program's weekly physical activity goal.
	Daily step count data for the 6-month study duration was exported from
	participants' personal Fitbit accounts into an Excel spreadsheet. Fitness
	was measured using the 6-Minute Walk Test (6-MWT), which measures
	the distance in feet that an individual can walk in six minutes. Change in
	fitness was calculated as the change in feet on the 6-MWT from baseline
	to 6-months. Weight was measured and reported as the change in body
	weight from baseline to 6-months.
	Statistical Analysis: Linear regression models were used to evaluate the
	association between average daily step count over the 6-month study
	duration and the pre-post changes in participants' weight and fitness.
	Penalized functional regression models were used to evaluate the time-
	varying association between daily step count collected and the pre-post
	changes in participants' weight and fitness. A p-value of 0.05 was
	considered statistically significant.
Data Collection	Timing of Measurements: Daily step count data was collected
Summary	continuously throughout the 6-month study duration. Body weight and
	fitness were measured at baseline and after 6-months.

	Dependent Variables: Changes in step count data according to the Fitbit
	Zip, changes in body weight and fitness after the 6-month intervention
	Independent Variables: Goal-setting component of the weight loss
	program, weekly group sessions led by lifestyle coaches targeting fitness
	and healthy eating, and several weekly text messages from research staff
	Control Variables: Body weight and fitness at baseline
	Initial: 43 ( Males Females)
	Attrition (final N): 34 (13 males and 21 females)
	Age: Mean age: $50.2$ years, $SD = 11.0$
	Ethnicity: Non-Hispanic white
	Other relevant demographics: Clinical characteristics: eight participants
	(23.5%) had a schizophrenia spectrum disorder, 17 participants (50.0%)
	had major depressive disorder, and nine participants (26.5%) had bipolar
	disorder. Education: two participants (5.9%) had less than a high school
	education, 11 participants (32.4%) attended high school, 13 participants
	(38.2%) attended some college, and eight participants (23.5%) earned a
	college degree. Living situation: 27 participants (79.4%) lived
Description of	independently, six participants (17.7%) lived with family, and one
Actual Data	participant (2.9%) lived in supported housing/assisted living. Marital
Sample	status: 11 participants (32.4%) were never married, three participants
	(8.8%) were currently married, and 20 participants (58.8%) were
	previously married. Employment/Insurance coverage: seven participants
	(20.6%) were currently employed (part or full-time), eight participants
	(23.5%) were enrolled in Medicaid only, eight participants (23.5%) were
	enrolled in Medicare only, and 18 participants (52.9%) were dual eligible
	(enrolled in Medicaid and Medicare). Six participants (17.7%) were
	current smokers.
	Anthropometrics: Mean weight: 231.9 pounds, SD = 46.7, mean BMI:
	38.5 kg/m2, SD = 9.3, fitness: 1303.8 feet in 6-MWT, SD = 323.2
	Location: Urban community mental health center in southern New
	Hampshire

	Key Findings: Due to rolling enrollment at the start of the study,
	participants had their Fitbit Zips to use for an average of $181.7$ days (SD =
	34.7), with a median of 181.5 days (interquartile range = $169$ to $196$ ).
	Participants wore their Fitbit Zips for a mean of $86.2\%$ (SD = $18.4\%$ ) of
	the days that they had the Fitbit Zip to use (median = $94.0\%$ of the days,
	interquartile range = $82.0\%$ to $97.0\%$ ). Participants achieved an average
	of 4453.5 (SD = $2707.4$ ) steps each day, with average daily step counts
	ranging from 1037.6 (SD = 767.9) steps to 11,366.3 (SD = 3416.9) steps.
	In total, 21 participants (61.8%) achieved 10,000 steps or more on at least
	one day during the study. These 21 participants achieved 10,000 daily
	steps or more for a mean of 16.1% (SD = $21.0\%$ ) of the days that they had
	the Fitbit Zip to use (median = 25.7% of the days, interquartile range =
	0.9% to 25.7%). Three participants (8.8%) achieved 10,000 daily steps or
	more on at least half of the days that they were enrolled in the study.
Summary of	There was a significant association between participants' average daily
Results	step count and weight loss. For every 1000 step increase in participants'
	daily average step count, they experienced a decrease in weight of 1.78
	pounds (F = 5.07; df = 1, 32; p = 0.0314). The relationship between
	average daily step count and change in fitness was not significant. If
	participants' average daily step count increased by 1000 steps, it
	corresponded to an increase of 18.79 feet on the 6-Minute Walk Test ( $F =$
	1.92; $df = 1, 31; p = 0.176$ ). In the penalized functional regression
	models, the time-varying relationship between daily step count and weight
	loss (permutation test statistic = $0.180$ ; p = $0.264$ ) and improved fitness
	(permutation test statistic = $0.076$ ; p = $0.574$ ) were not significant. This
	suggests that there was no specific period of time for which an increase in
	steps was significantly associated with either weight loss or improved
	fitness.
	Other Findings:

	At 6-months, higher average daily step count was associated with greater
	weight loss, but not improved fitness. These findings suggest that
	wearable devices and their associated smartphone applications may serve
	as valuable tools for supporting community-based weight loss efforts for
	people with serious mental illness. Importantly, it was observed that a
	higher average daily step count over the 6-month program duration was
Author	associated with greater weight loss. Therefore, it appears that providing
Conclusion	participants with serious mental illness the recommendation to collect
	more steps each day and maintain a high average daily step count
	throughout participation in a lifestyle intervention may contribute to
	greater weight loss. These are preliminary findings and should be
	interpreted with caution, but they offer promise regarding the potential
	benefits of using wearable devices to support lifestyle interventions
	delivered through community mental health settings.
	Strengths: long study duration, highly engaged participants, Fitbit Zips
	were used to support self-monitoring, goal-setting, and tracking progress
	over time
	Limitations: small sample size, predominantly female sample, lacking
	racial or ethnic diversity, relatively high dropout rate, blinding was not
Reviewer	utilized, analyses were based on the participants who completed the 6-
Comments	month intervention, because Fitbit Zips were integrated as part of the
	behavioral weight loss program, the results cannot separate the impact of
	group education and support from use of the wearable device, and all
	participants were receiving services through community mental health
	settings, thus findings are likely not representative of individuals with
	serious mental illness not currently receiving services
	This study was supported by the National Institute of Mental Health (R01
	MH089811-01) and the United States Centers for Disease Control and
Funding Source	Prevention Health Promotion and Disease Prevention Research Center
	(Cooperative Agreement Number U48DP005018). The funders had no

role in study design, data collection and analysis, decision to publish, or
preparation of the manuscript. The authors report no conflicting interests.

## **Quality Criteria Checklist: Primary Research**

Symbols Used	Explanation
	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
0	exceptionally week

Select a rating from the drop-down menu  $\checkmark$ 

Re	Relevance Questions		
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Yes
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	Yes
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

# If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Va	Validity Questions				
1.	Was the <u>research question</u> clearly stated?	1	Yes		
	1.1. Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	Yes		
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes		
	1.3. Were the target population and setting specified?	1.3	Yes		
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	Yes		
	progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Yes		
	without omitting criteria critical to the study? 2.2. Were criteria applied equally to all study groups?	2.2	N/A		
	2.3. Were health, demographics, and other characteristics of subjects described?	2.3	Yes		
	population?	2.4	Unclear		
3.	Were study groups comparable?	2	Var		
	3.1. Was the method of assigning subjects/patients to groups described and	3	res		
	unbiased? (Method of randomization identified if RCT) 3.2. Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	N/A		
		3.2	Yes		

	3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Yes
	3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.4	N/A
	3.5.	If case control study, were potential confounding factors comparable for cases and controls? (If case series or trial with subjects serving as own control, this	3.5	N/A
	3.6.	criterion is not applicable. Criterion may not be applicable in some cross- sectional studies.) If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g., "gold standard")?	3.6	N/A
	4. Was r	nethod of handling withdrawals described?	4	Yes
	4.1.	Were follow up methods described and the same for all groups?	4.1	N/A
	4.2.	Was the number, characteristics of withdrawais (i.e., dropouts, lost to follow	4.1	IN/A
		each group? (Follow up goal for a strong study is 80%.)	4.2	Yes
	4.3.	Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	No
	4.4.	Were reasons for withdrawals similar across groups	4.4	Yes
	4.5.	If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4.5	N/A
		linding used to prevent introduction of bios2	4.3	IN/A
	5. was <u>c</u>	Inding used to prevent introduction of blas?	5	No
	5.1.	blinded to treatment group, as appropriate?		) I
	5.2.	Were data collectors blinded for outcomes assessment? (If outcome is	5.1	No
		measured using an objective test, such as a lab value, this criterion is assumed	5.2	No
	53	to be met.)		
	5.5.	risk factors blinded?	5.3	N/A
	5.4.	In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4	N/A
	5.5.	In diagnostic study, were test results blinded to patient history and other test results?	5.5	N/A
	6. Were	intervention/therapeutic regimens/exposure factor or procedure and any	6	Yes
	comp	arison(s) described in detail? Were <u>intervening factors</u> described?	6.1	Vec
	0.1.	studied?	0.1	105
	6.2.	In observational study, were interventions, study settings, and	6.2	N/A
	<b>C</b> 2	clinicians/provider described?	6.3	Yes
	0.3.	to produce a meaningful effect?	6.4	Yes
	6.4.	Was the amount of exposure and, if relevant, subject/patient compliance		
	6.5	measured?	6.5	Yes
	6.5. 6.6	Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.6	N/A
	6.7.	Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?	6.7	N/A
	6.8.	In diagnostic study, were details of test administration and replication	0.7	
		sufficient?	6.8	N/A
	7. Were	outcomes clearly defined and the measurements valid and reliable?	7	Yes
	7.1.	were primary and secondary endpoints described and relevant to the question?	7.1	Yes
	7.2.	Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A
_				

	7.3.	Was the period of follow-up long enough for important outcome(s) to occur?	7.3	N/A
	7.4.	Were the observations and measurements based on standard, valid, and reliable data collection instruments /tests /procedures?	7.4	Yes
	7.5.	Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.6.	Were other factors accounted for (measured) that could affect outcomes?	7.6	Yes
	7.7.	Were the measurements conducted consistently across groups?	7.7	N/A
8.	Was t	ne <u>statistical analysis</u> appropriate for the study design and type of outcome	8	Yes
	indica	tors?		
	8.1.	Were statistical analyses adequately described the results reported	8.1	Yes
	8.2.	Were correct statistical tests used and assumptions of test not violated?	8.2	Yes
	8.3.	Were statistics reported with levels of significance and/or confidence intervals?	8.3	Yes
	8.4.	an analysis of outcomes for those maximally exposed or a dose-response	8.4	No
	8.5.	Were adequate adjustments made for effects of confounding factors that	8.5	No
	8.6.	might have affected the outcomes (e.g., multivariate analyses)? Was clinical significance as well as statistical significance reported?	8.6	Yes
	8.7.	If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A
9.	Are <u>co</u>	nclusions supported by results with biases and limitations taken into	9	Yes
	consid	eration?	0.1	Vec
	9.1.	Is there a discussion of findings?	7.1	105
	9.2.	Are biases and study limitations identified and discussed?	9.2	Yes
10.	Is bias	due to study's <u>funding or sponsorship</u> unlikely?	10	Yes
	10.1	. Were sources of funding and investigators' affiliations described?	10.1	Yes
	10.2	. Was there no apparent conflict of interest?	10.2	Yes
MI	NUS/N	EGATIVE (-)		

If most (six or more) of the answers to the above validity questions are "No," the report should be designated with a minus (-) symbol on the Evidence Worksheet.

## NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

## PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format as found in JADA.	Rosenberger, M., Buman, M., Haskell, W., McConnell, M., & Carstensen, L. (2016). 24 Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. <i>Medicine and Science in Sports and Exercise</i> , <i>48</i> (3), 457–465.
Study design: Use algorithm –	Validity study
RCT, cohort, etc	
Study Class (A,B,C,D)	С
Research Quality Rating	NEUTRAL (Ø)
This rating tells if the research	
design is good (+), bad (-) or	
neutral (Ø)	
This is determined by the quality	
criteria list. Delete the ratings	
that do not apply (i.e. if positive,	
delete minus/negative and	
neutral).	
	rurpose/ropulation Stualea/Practice Studied
Research purpose: What is the	To compare the output from commercially available wearable devices using current
research question being	standards for objective measurement of sleep, sedentary behaviors (SED), light-
investigated in the study?)	intensity physical activity (LPA), moderate-to-vigorous physical activity (MVPA),
	determine the best ways to measure the full 24 hours of activity behavior to guide
	future clinical studies and recommendations
Inclusion criteria: requirements	Unclear
for study eligibility	
Exclusion criteria (conditions	Unclear
that make individual ineligible)	
Recruitment	The Stanford University community and surrounding areas through word-of-mouth
	with an effort to include equal numbers of men and women over a wide age range
Blinding used: some of the	NA
from knowing cartain information	
that might lead to conscious or	
unconscious bias on their part	
invalidating the results	
Description of study protocol	Participants wore nine devices for 24-hours: Actigraph GT3X+, activPAL, Fitbit
What happened in the study?	One, GENEactiv, Jawbone Up, LUMOback, Nike Fuelband, Omron pedometer,
	and Z-Machine. Comparisons (to standards) were made for total sleep time (Z-
	machine), time spent in SED (activPAL), LPA duration (GT3x+), MVPA duration
	(GT3x+), and total steps per day (Omron).
Intervention: Describe	Participants came to the laboratory where height, weight, age, and gender were
interventions, regimens, risk	collected and recorded. Software was used to submit participant-specific
<i>Jactors, or procedures studied.</i>	Information to each device for initialization and calibration. Participants also
	when to put on the devices and how to wear them. Dorticipants were asked to wear
	all nine devices for one full day of activity and one full night of sleep. Devices
	were worn from approximately the time a participant woke up until the participant
	woke up the next morning. Device feedback was not provided to the participant
	except in cases where the data was presented on the device itself. No interventions
	were introduced such as step goals, vibrations to interrupt sedentary behavior, or
	other guidelines for the participant. Device data were downloaded after the
	participant returned the study kit. Participants could view their data after the

	conclusion of their participation if they were willing to stay through data download. No written reports were provided to the participant. Data were either downloaded to the computer (Fitbit, GT3X+, Fuelband, and activPAL) or through the phone application (LUMOback and Jawbone) for devices that lack desktop software. Devices compared to the Z-machine for measuring sleep duration included the Fitbit, Jawbone, GENEactiv, and GT3X+. Devices compared to the activPAL for measuring SED duration included the GT3X+, GENEactiv, LUMOback, and Fitbit. Devices compared to the GT3X+ for measuring LPA duration included the Fitbit and GENEactiv. Devices compared to the GT3X+ for measuring MVPA duration included the Jawbone, Fitbit, GENEactiv, and Fuelband. Devices compared to the Omron for measuring total steps included the Jawbone, Fitbit, Fuelband, GT3X+, LUMOback, and activPAL.
Statistical analysis: List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Standard sample calculations were conducted to set goals for recruitment, and alpha was set at .05 with the confidence interval set to 95%. Separate sample calculations were conducted for each domain. Statistical analyses, including mean absolute percent error (MAPE), equivalence testing, and Bland-Altman plots were performed to determine statistically significant differences as well as agreement
<b>Timing of measurements:</b> when outcomes were measured; usually baseline and one or more later times	among devices. Height, weight, age, and gender were collected and recorded at baseline. Sleep, SED, LPA, MVPA, and steps were collected continuously throughout the 24-hour intervention. Data were downloaded after the 24-hour intervention.
<b>Dependent variables:</b> <i>outcomes</i> <i>that are measured or registered;</i> <i>variable whose change or</i> <i>different states the researcher</i> <i>wants to understand, explain, or</i> <i>predict</i>	Total sleep time according to Fitbit, Jawbone, GENEactiv, and GT3X+; time spent in SED according to GT3X+, GENEactiv, LUMOback, and Fitbit; LPA duration according to Fitbit and GENEactiv; MVPA duration according to Jawbone, Fitbit, GENEactiv, and Fuelband; and total steps per day according to Jawbone, Fitbit, Fuelband, GT3X+, LUMOback, and activPAL
<b>Independent variables</b> (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Free living conditions (e.g. participants' daily obligations, lifestyles, level of physical fitness, stress levels)
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	Total sleep time according to Z-machine; time spent in SED according to activPAL; LPA and MVPA duration according to GT3x+; and total steps per day according to Omron
Initial n (e.g. 731 (298 males, 433 females)) Record number that entered study – not the number screened.	40 participants (19 males and 21 females)
<b>Final n</b> (attrition) number of subjects that completed study	40 participants (19 males and 21 females)
Age usually mean or range	21 to 76 years
Ethnicity (if given)	Unclear
<b>Other relevant demographics:</b> <i>demographics describe the</i> <i>population (students, athletes, etc)</i>	NA
Anthropometrics: e.g. were groups same or different on	NA

	1
<i>important physical measures</i> <i>(BMI, fitness level)</i>	
<b>Location:</b> <i>Where did the study</i>	Stanford University community and surrounding areas (California)
take place? City or country	
Summary of Results: Abstract	Mean error analyses for the devices ranged from 8.1% for GT3X+ to 16.9% for
results including quantitative data	GENEactiv when measuring sleep duration; 9.5% for LUMOback to 65.8% for
and statistics. Include statistical	GENEactiv when measuring SED; 19.7% for GENEactiv to 28.0% for Fitbit when
significance: P-values, confidence	measuring LPA; 51.8% from Jawbone to 92.0% from Fuelband when measuring
intervals (CI), relative risk (RR),	MVPA; and 14.1% from GT3X+ to 29.9% from Fuelband when measuring total
odds ratios (OR), likelihood ratio,	steps per day. Equivalence analyses indicated only two comparison devices were
number needed to treat, power	significantly equivalent to standards: $GI3X+$ for sleep (90% CI), and LUMOback for SED (00% CI). Pland Altman plats had mean differences ranging from 4
analysis if available.	101 SED (90% C1). Bland-Altman plots had mean differences fanging from 4 minutes for $GT3Y + to 36$ minutes for Eithit and GENEactive when measuring sleep
	duration: 18 minutes for LUMOback to 162 minutes for GENEactiv when
	measuring SED: 43 minutes for GENEactiv to 64 minutes for Fifthit when
	measuring LPA: 48 minutes for Jawbone to 598 minutes for Fuelband when
	measuring MVPA; and 698 steps for GT3X+ to 2258 steps for activPAL when
	measuring total steps per day.
	Author's Conclusions
Author conclusion: <i>paraphrase</i>	Findings suggest that measurement of activity domains (sleep, sedentary behavior,
that stated by study author in	and physical activity) is highly varied among wearable devices when tested outside
body of the report or abstract	of the laboratory. While this may sound discouraging, the ability to measure very
	specific behaviors has greatly increased with the introduction of a large number of
	the predictable error that comes from comparing actigraphy to polycompography
	For steps, many of the devices were different from the standard, but gave similar
	results to each other implying some predictable agreement among devices
	Currently, 24-hour activity measurement is only possible with research-grade
	devices. None of the commercial wearable devices provide all the measures of the
	24-hour model. The future of activity measurement should aim for accurate 24-
	hour measurement as a goal. Researchers should continue to select measurement
	devices based on their primary outcomes of interest. Evaluation of devices will be
	an ongoing area of research because of the rapid changes in wearable technology.
Reviewer comments: Note	Strengths: zero percent attrition, the use of numerous consumer and reference
strengths and limitations of study;	devices, testing the devices in a free-living environment as they are designed for,
validity and generalizability	Limitations: inclusion and exclusion criteria and participant demographics were
vour comments should be	not discussed blinding was not utilized standards were based on common field-
italicized)	based measures, not gold standards used in the laboratory, therefore, both the test
	device and criterion device introduce substantial error into the comparisons,
	placement of activity monitors can affect how well these devices match up to
	standards, and the functions of these devices change with every software and
	hardware update, therefore, not every possible update can be evaluated with the
	research at one particular point in time
	Funding source: Grant R3/-AG008816 from the National Institute on Aging to
	Laura L. Carstensen. Dr. Kosenberger was a postdoctoral fellow supported on the
	research support from Apple Inc. The results of this study do not constitute
	endorsement by the American College of Sports Medicine. The authors have no
	potential conflicts-of-interest to disclose.

Table 3.2.a. Quality Criteria Checklist: Primary Research

RELEVANCE OUESTIONS					
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)		Y	N	U	N
Rosenberger, M., Buman, M., Haskell, W., McConnell, M., & Carstensen, L. (2016). 24 Hours		E S	0	C N	А
of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Medicine				L E	
and Science in Sports and Exercise, 48(3), 457–465. doi:10.1249/MSS.000000000000778				A	
1 Would implementing the studied intervention or procedure (if found successful) result in	1			X	+
improved outcomes for the nations/clients/nonulation group? (Not Applicable for some	1			11	
enidemiological studies)					
2 Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	x			+
nonulation group would care about?	2				
3 Is the focus of the intervention or procedure (independent variable) or topic of study a	3		x		+
common issue of concern to dietetics practice?	5		11		
4 Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	x			
If the answers to all of the above relevance questions are "ves" the report is eligible for design	nation w	$\frac{1}{1}$	nlu	s (+	) on
the Evidence Quality Worksheet depending on answers to the following validity questions	iunon n	ann a	pin	5 ( '	) 011
WALIDITY OUESTIONS					
VALIDITY QUESTIONS		Y	r	N	UN
1. was the <u>research question</u> clearly stated?		E		0	N A
This is usually stated at end of the introduction and just before methods section.		5		1	ĭ
			,	1	E A
		2	<b>`</b>	1	R
1.1 Was the specific intervention(s) or procedure (independent variable(s)) identified?	1.1	2	Κ		
This is often called the treatment and explained in the methods section.					
1.2 Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Σ	Κ		
These are sometimes called the endpoints; the results section reports the outcomes, but					
this information should be in the methods section, too					
1.3 Were the target population and setting specified?	1.3				Х
The target population is group for whom findings may be applicable; look for this in the					
introduction and in the methods section					
2. Was the <u>selection</u> of study subjects/patients free from bias?		Y E		N I O I	U N N A
		s			C
				1	Ē
					A R
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1			Х	
diagnostic or prognosis criteria), and with sufficient detail and without omitting					
criteria critical to the study?					
The authors should give several points about the inclusion/exclusion criteria such as the					
age range of the subjects, disease condition (like hyperlipidemia) required for					
inclusion. Exclusion criteria should be listed, too, although some are					
understood. For example if the ages for inclusion are 18 to 70, the authors will					
probably not specifically note that children and people over age 70 were					
excluded. Most of the time, however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD).					
2.2 Were criteria applied equally to all study groups?	2.2				X
2.3 Were health, demographics, and other characteristics of subjects described?	2.3			Х	
There is usually a Table 1 summarizing demographics and characteristics at baseline.					
Groups are <u>not</u> different if the P-Value is $> 0.05$ . If there has been a previous					
paper describing the study population, that paper may be referenced and you					
would need to go back to the original publication to see that Table 1.	1				$\perp$
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4				Х
The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may					
only say that the patients came from the same clinic from people who met the inclusion criteria					1

3. W	/ere <u>study groups comparable</u> ?		Y	N O	U N	N A
	There is usually a Table 1 summarizing demographics and characteristics at baseline.		ŝ	Ŭ	C	
	<i>Groups are</i> <u>not</u> different if the <i>P</i> -Value is $> 0.05$ .				E	
					A R	Х
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone					
	disliked their supplement, they were allowed to change groups – this is not unbiased!					
	3.2 Were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2				Х
	demographics) similar across study groups at baseline? See Table I for this - there					
	should be no significant differences across study groups in an intervention study.					
	3.3 Were concurrent controls used? (Concurrent preferred over historical controls.)	3.3	Х			
	Most RCTs use a concurrent control group. Occasionally an intervention study will use a					
	prior study as a control group; that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group. A crossover study					
	where the subject acts as his/her own control is use of concurrent control.					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4				Х
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	description and results sections.					
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	Х			
	reference standard (e.g. "gold standard")?					
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard). In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Benedict) to predict energy needs, a gold standard measure of REE (Indirect					
	Calorimetry) is used.					
4.	Was method of handling withdrawals described?		Y	N O	UN	N A
			S	Ŭ	C	л
					E	
					AR	Х
	4.1 Were follow up methods described and the same for all groups?	4.1				Х
	4.2 Was the number characteristics of withdrawals (i.e. dropouts lost to follow up	42				X
	attrition rate) and/or response rate (cross-sectional studies) described for each group?					
	(Follow up goal for a strong study is 80 %.)					
	This should be found in the results section. If there is attrition $> 20\%$ it is important to					
	note that on the worksheet (as a note in the results section or in the reviewer					
	comments at the very bottom)					
	4.3 Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	X			
	This information is often presented in a figure with # recruited. # enrolled (this is initial					
	N), # remaining at end of study period (final N). Sometimes the reasons that subjects					
	withdrew or were dropped is given in the figure or in the text (results section).					
	4.4 Were reasons for withdrawals similar across groups?	4.4	1			Х
	If there is a large attrition from one group and not others. vou would want to look for a					-
	reason why; the answer to this question would then be no.					
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of	4.5	Х			
----	-----------------------------------------------------------------------------------------------	-----	---	--------	--------	--------
	test under study?					I
	The test under study should be compared to reference test all the time. An example of this					I
	might be using a DEXA machine to measure percent body fat only if a subject's					1
	BMI was $> 35$ but bioimpedance analyzer indicated body fat $< 30\%$ .					1
5.	Was <u>blinding</u> used to prevent introduction of bias?		Y	N	UN	N
			S	U	C	А
					L E	I
					A	1
					X	1
	5.1 In intervention study were subjects clinicians/practitioners and investigators blinded	51				X
	to treatment group as appropriate?	0.1				
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					1
	studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					I
	an RCT. but obviously, the subjects and practitioners knew who was getting MNT					1
	and who was not. Therefore, you would not answer question 5.1 NO. It was					1
	appropriate for the dietitians and patients to know they were receiving MNT.					1
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2			Х	
	using an objective test, such as a lab value, this criterion is assumed to be met.)					1
	Answer ves, if a lab test was used to measure an outcome. A method of blinding a diet					1
	study is to have separate people analyzing the data (not the same ones who were					1
	collecting the data).					1
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					I
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					I
	study is to have separate people analyzing the data (not the same ones who were					1
	collecting the data).					I
	5.4 In case control study, was case definition explicit and case ascertainment not	5.4				Х
	influenced by exposure status?					I
	Establish who is a case and who is a control at the beginning of the study.					1
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5			Х	
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any		Y	N O	U N	N A
	comparison(s) described in detail? Were intervening factors described?		ŝ	Ŭ	Ĉ	
					L E	1
			Х		A R	1
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1				Х
	6.2. In observational study were interventions study settings and clinicians/provider	62	X			
	described?	0				1
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to	6.3	X			
	produce a meaningful effect?					I
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					1
	difference in lab values for cholesterol; however, 12 days would not be long					1
	enough)					1
	6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.4			Х	
	How long did the treatment last? Did the patient follow directions?					1
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5				Х
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)					I
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					1
	discussion section. It is likely there were no unplanned treatments, so a "no"					1
1	answer is not a problem overall.					1
	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups?	6.7				Х
	For a study to be valid and unbiased, it is important that this be yes.					1
	6.8 In diagnostic study, were details of test administration and replication sufficient?	6.8	Х			
	Usually answer n/a for diet study.					L

7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E	N A
			Х		A R	
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X			
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT; Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures? Check that surveys were validated	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? <i>Precision is reproducibility or repeatability.</i>	7.5	X			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S	N O	U N C L E A	N A
		0.1	A		R	
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	X			
	8.2 Were correct statistical tests used and assumptions of test not violated? You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P</i> -value) and/or confidence intervals (mean $\pm$ CI)	8.3	Х			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4				X
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5				Х
	<ul> <li>8.6 Was clinical significance as well as statistical significance reported?</li> <li>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical significance (P-value) and clinical significance (compare to standard of &lt; 200 mg/do for normal cholesterol). A problem can occur when only statistical</li> </ul>	8.6	X			

significance is reported. Reducing cholesterol from 300 to 250 might be statistic significant, but clinically it is still abnormal.	cally				
8.7 If negative findings, was a power calculation reported to address type 2 error? <i>Type II (<math>\beta</math> error is a false negative that happens when the investigators fail to reject</i> <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of $n=xx$ is needed to provide 80% power."	8.7 <i>the</i>				X
9. Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S X	N O	U N C L E A	N A
0.1 Is there a discussion of findings?	0.1	v		R	
4.1 is there a discussion of findings?	9.1	Λ			
0.2 Are biases and study limitations identified and discussed?	0.2	v			
This will be in the discussion of finding section that follows the results	9.2	Λ			
10 Is bias due to study's funding or sponsorship unlikely?		Y	N	U	N
Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		E S X	0	N C L A R	Α
<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biase</li> </ul>	10.1 d.	X			
10.2 Was there no apparent conflict of interest? If an investigator is testing a piece of equipment, process or drug that s/he developed, could be a conflict of interest.	<i>it</i> 10.2	X			
SYMBOL			<u>.                                    </u>		
MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report sh minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (0)	nould be design	nated	with	h a	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Academy of Nutrition and Dietetics Evidence Analysis Library® Worksheet Template and Quality Criteria Checklist: Primary Research

Citation	Rosenberger, M., Buman, M., Haskell, W., McConnell, M., & Carstensen, L. (2016). 24 Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Medicine and Science in Sports and Exercise, 48(3), 457–465. doi:10.1249/MSS.000000000000778
Study Design	Validity study
Class	C
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\boxtimes \otimes$ (Neutral)
Research Purpose	To compare the output from commercially available wearable devices using current standards for objective measurement of sleep, sedentary behaviors (SED), light-intensity physical activity (LPA), moderate-to- vigorous physical activity (MVPA), and steps in a free-living environment. The ultimate goal of this research is to determine the best ways to measure the full 24 hours of activity behavior to guide future clinical studies and recommendations.
Inclusion Criteria	Unclear
Exclusion Criteria	Unclear
Description of Study Protocol	<ul> <li>Recruitment: The Stanford Oniversity community and surrounding areas</li> <li>through word-of-mouth with an effort to include equal numbers of men</li> <li>and women over a wide age range</li> <li>Design: Participants wore nine devices for 24-hours: Actigraph GT3X+,</li> <li>activPAL, Fitbit One, GENEactiv, Jawbone Up, LUMOback, Nike</li> <li>Fuelband, Omron pedometer, and Z-Machine. Comparisons (to</li> <li>standards) were made for total sleep time (Z-machine), time spent in SED</li> <li>(activPAL), LPA duration (GT3x+), MVPA duration (GT3x+), and total</li> <li>steps per day (Omron).</li> <li>Blinding used (if applicable): NA</li> <li>Intervention (if applicable): Participants came to the laboratory where</li> <li>height, weight, age, and gender were collected and recorded. Software</li> <li>was used to submit participant-specific information to each device for</li> <li>initialization and calibration. Participants also received a study kit</li> <li>including device supplies, written, and oral instructions of when to put on</li> <li>the devices and how to wear them. Participants were asked to wear all</li> </ul>

	nine devices for one full day of activity and one full night of sleen
	The devices for one funday of activity and one funding of sleep.
	Devices were worn from approximately the time a participant woke up
	until the participant woke up the next morning. Device feedback was not
	provided to the participant except in cases where the data was presented
	on the device itself. No interventions were introduced such as step goals,
	vibrations to interrupt sedentary behavior, or other guidelines for the
	participant. Device data were downloaded after the participant returned
	the study kit. Participants could view their data after the conclusion of
	their participation if they were willing to stay through data download. No
	written reports were provided to the participant. Data were either
	downloaded to the computer (Fitbit, GT3X+, Fuelband, and activPAL) or
	through the phone application (LUMOback and Jawbone) for devices that
	lack desktop software. Devices compared to the Z-machine for measuring
	sleep duration included the Fitbit, Jawbone, GENEactiv, and GT3X+.
	Devices compared to the activPAL for measuring SED duration included
	the GT3X+, GENEactiv, LUMOback, and Fitbit. Devices compared to
	the GT3X+ for measuring LPA duration included the Fitbit and
	GENEactiv. Devices compared to the GT3X+ for measuring MVPA
	duration included the Jawbone, Fitbit, GENEactiv, and Fuelband.
	Devices compared to the Omron for measuring total steps included the
	Jawbone, Fitbit, Fuelband, GT3X+, LUMOback, and activPAL.
	Statistical Analysis: Standard sample calculations were conducted to set
	goals for recruitment, and alpha was set at .05 with the confidence interval
	set to 95%. Separate sample calculations were conducted for each
	domain. Statistical analyses, including mean absolute percent error
	(MAPE), equivalence testing, and Bland-Altman plots were performed to
	determine statistically significant differences as well as agreement among
	devices
Data Collection	Timing of Massuraments: Height weight ago and conder ware collected
Summary	Thinning of Measurements. Height, weight, age, and gender were collected
	and recorded at baseline. Sleep, SED, LPA, MVPA, and steps were

	collected continuously throughout the 24-hour intervention. Data were				
	downloaded after the 24-hour intervention.				
	Dependent Variables: Total sleep time according to Fitbit, Jawbone,				
	GENEactiv, and GT3X+; time spent in SED according to GT3X+,				
GENEactiv, LUMOback, and Fitbit; LPA duration according to					
	GENEactiv; MVPA duration according to Jawbone, Fitbit, GENEactiv,				
	and Fuelband; and total steps per day according to Jawbone, Fitbit,				
	Fuelband, GT3X+, LUMOback, and activPAL				
	Independent Variables: Free living conditions (e.g. participants' daily				
	obligations, lifestyles, level of physical fitness, stress levels)				
	Control Variables: Total sleep time according to Z-machine; time spent in				
	SED according to activPAL; LPA and MVPA duration according to				
	GT3x+; and total steps per day according to Omron				
	Initial: 40 (19 Males 21 Females)				
	Attrition (final N): 40				
	Age: 21-76 years				
Description of	Ethnicity: Unclear				
Actual Data Sample	Other relevant demographics: NA				
	Anthropometrics: NA				
	Location: Stanford University community and surrounding areas				
	(California)				
	Key Findings: Mean error analyses for the devices ranged from 8.1% for				
	GT3X+ to 16.9% for GENEactiv when measuring sleep duration; 9.5%				
	for LUMOback to 65.8% for GENEactiv when measuring SED; 19.7%				
	for GENEactiv to 28.0% for Fitbit when measuring LPA; 51.8% from				
Summary of	Jawbone to 92.0% from Fuelband when measuring MVPA; and 14.1%				
Results	from GT3X+ to 29.9% from Fuelband when measuring total steps per				
	day. Equivalence analyses indicated only two comparison devices were				
	significantly equivalent to standards: GT3X+ for sleep (90% CI), and				
	LUMOback for SED (90% CI). Bland-Altman plots had mean differences				
	ranging from 4 minutes for GT3X+ to 36 minutes for Fitbit and				

GENEactiv when measuring sleep duration; 18 minutes for LUMO			
	162 minutes for GENEactiv when measuring SED; 43 minutes for		
	GENEactiv to 64 minutes for Fitbit when measuring LPA; 48 minutes for		
	Jawbone to 598 minutes for Fuelband when measuring MVPA; and 698		
	steps for GT3X+ to 2258 steps for activPAL when measuring total steps		
	per day.		
	Other Findings:		
	Findings suggest that measurement of activity domains (sleep, sedentary		
	behavior, and physical activity) is highly varied among wearable devices		
	when tested outside of the laboratory. While this may sound		
	discouraging, the ability to measure very specific behaviors has greatly		
	increased with the introduction of a large number of wearable devices.		
	For sleep, many of the devices can measure total sleep time with the		
	predictable error that comes from comparing actigraphy to		
	polysomnography. For steps, many of the devices were different from the		
Author	standard, but gave similar results to each other, implying some predictable		
Conclusion	agreement among devices. Currently, 24-hour activity measurement is		
	only possible with research-grade devices. None of the commercial		
	wearable devices provide all the measures of the 24-hour model. The		
	future of activity measurement should aim for accurate 24-hour		
	measurement as a goal. Researchers should continue to select		
	measurement devices based on their primary outcomes of interest.		
	Evaluation of devices will be an ongoing area of research because of the		
	rapid changes in wearable technology.		
	Strengths: zero percent attrition, the use of numerous consumer and		
	reference devices, testing the devices in a free-living environment as they		
Reviewer	are designed for, and examining several different activity domains		
Comments	collected by the devices		
	Limitations: inclusion and exclusion criteria and participant		

	were based on common field-based measures, not gold standards used in
	the laboratory, therefore, both the test device and criterion device
	introduce substantial error into the comparisons, placement of activity
	monitors can affect how well these devices match up to standards, and the
	functions of these devices change with every software and hardware
	update, therefore, not every possible update can be evaluated with the
	research at one particular point in time
	Grant R37-AG008816 from the National Institute on Aging to Laura L.
	Carstensen. Dr. Rosenberger was a postdoctoral fellow supported on the
	same grant. Stanford Cardiovascular Medicine has received in-kind
Funding Source	mobile health research support from Apple Inc. The results of this study
	do not constitute endorsement by the American College of Sports
	Medicine. The authors have no potential conflicts-of-interest to disclose.

# Quality Criteria Checklist: Primary Research

Symbols Used	Explanation
4	<b>Positive</b> – Indicates that the report has clearly addressed issues of
Т	inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor
0	exceptionally week

Select a rating from the drop-down menu ↓

Re	Relevance Questions				
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear		
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes		
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No		
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes		

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Validity Questions				
1.	Was the <u>research question</u> clearly stated?	1	Yes	
		1.1	Yes	

	1.1. Was the specific intervention(s) or procedure (independent variable(s))	1.2	Yes
	identified?		
	1.2. Was the outcome(s) (dependent variable(s)) clearly indicated? 1.3. Were the target population and setting specified?	1.3	Unclear
2.	Was the <u>selection</u> of study subjects/patients free from bias?	2	No
	2.1. Were inclusion/exclusion criteria specified (e.g., risk, point in disease		
	progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	No
	Without omitting criteria critical to the study?	2.2	N/A
	2.3. Were health, demographics, and other characteristics of subjects described?	2.2	No
	2.4. Were the subjects/patients a representative sample of the relevant	2.5	INO
	population?	2.4	Unclear
3.	Were study groups comparable?	3	N/A
	3.1. Was the method of assigning subjects/patients to groups described and	5	14/21
	Unblased? (Method of randomization identified if RCI)	3.1	N/A
	demographics) similar across study groups at baseline?	5.1	11/11
	3.3. Were concurrent controls used? (Concurrent preferred over historical	3.2	N/A
	controls.)		1.1/11
	3.4. If cohort study or cross-sectional study, were groups comparable on important	3.3	Yes
	contounding factors and/or were preexisting differences accounted for by usin appropriate adjustments in statistical analysis?	8	
	3.5. If case control study, were potential confounding factors comparable for cases	3.4	N/A
	and controls? (If case series or trial with subjects serving as own control, this		
	criterion is not applicable. Criterion may not be applicable in some cross-	3.5	N/A
	sectional studies.)		
	3.6. If diagnostic test, was there an independent blind comparison with an appropriate reference standard (e.g. "gold standard")?	3.6	Yes
4	Was method of handling withdrawals described?		
	4.1. Were follow up methods described and the same for all groups?	4	N/A
	4.2. Was the number, characteristics of withdrawals (i.e., dropouts, lost to follow	4.1	N/A
	up, attrition rate) and/or response rate (cross-sectional studies) described for	4.2	N/A
	each group? (Follow up goal for a strong study is 80%.)	1.2	X7
	4.3. Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Yes
	4.4. Were reasons for withdrawars similar across groups 4.5. If diagnostic test, was decision to perform reference test not dependent on	4.4	N/A
	results of test under study?	4.5	Yes
5.	Was blinding used to prevent introduction of bias?	-	TT 1
	5.1. In intervention study, were subjects, clinicians/practitioners, and investigators	5	Unclear
	blinded to treatment group, as appropriate?	5.1	N/A
	5.2. Were data collectors blinded for outcomes assessment? (If outcome is		
	to be met )	5.2	Unclear
	5.3. In cohort study or cross-sectional study, were measurements of outcomes and		
	risk factors blinded?	5.3	N/A
	5.4. In case control study, was case definition explicit and case ascertainment not	5.4	N/A
	influenced by exposure status?	5.4	
	5.5. In diagnostic study, were test results blinded to patient history and other test results?	5.5	Unclear
			X7
1		6	Yes

6.	5. Were intervention/therapeutic regimens/exposure factor or procedure and any 6.1 N/A					
	comparison(s) described in detail? Were <u>intervening factors</u> described?		1.0/1.1			
	6.1. In RCT or other intervention trial, were protocols described for all regimens studied?	6.2	Yes			
	6.2. In observational study, were interventions, study settings, and	6.3	Yes			
	clinicians/provider described?	6.4	Unclear			
	6.3. Was the intensity and duration of the intervention or exposure factor sufficient	0.1				
	to produce a meaningful effect?	6.5	N/A			
	measured?	6.6	N/A			
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) described?	6.7	N/A			
	6.7 Was the information for 6.4, 6.5, and 6.6 assessed the same way for all groups?					
	6.8. In diagnostic study, were details of test administration and replication	6.8	Yes			
7	Sufficients Were outcomes clearly defined and the measurements valid and reliable?	7	V			
<i>'</i> .	7.1. Were primary and secondary endpoints described and relevant to the	/	Yes			
	question?	7.1	Yes			
	7.2. Were nutrition measures appropriate to question and outcomes of concern?	7.2	N/A			
	7.3. Was the period of follow-up long enough for important outcome(s) to occur?	7.3	N/A			
	reliable data collection instruments/tests/procedures?	7.4	Yes			
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes			
	7.6. Were other factors accounted for (measured) that could affect outcomes?	7.6	Yes			
	7.7. Were the measurements conducted consistently across groups?	7.7	N/A			
8.	Was the statistical analysis appropriate for the study design and type of outcome	8	Yes			
	indicators?	0.1	V			
	appropriately?	8.1	Yes			
	8.2. Were correct statistical tests used and assumptions of test not violated?	8.2	Yes			
	8.3. Were statistics reported with levels of significance and/or confidence intervals?	8.3	Yes			
	an analysis of outcomes for those maximally exposed or a dose-response	8.4	N/A			
	analysis)? 8.5. Were adequate adjustments made for effects of confounding factors that	8.5	N/A			
	might have affected the outcomes (e.g., multivariate analyses)?	8.6	Yes			
	8.7. If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A			
9.	Are conclusions supported by results with biases and limitations taken into	9	Yes			
	consideration?	9.1	Yes			
	9.1. Is there a discussion of findings?	9.2	Ves			
10	9.2. Are biases and study limitations identified and discussed?	10	Vez			
10.	10.1. Were sources of funding and investigators' affiliations described?	10	Yes			
	10.2. Was there no apparent conflict of interest?	10.1	Yes			
1.41		10.2	res			
I If $m$	INUS/INEGATIVE (-) nost (six or more) of the answers to the above validity questions are "No," the report should numbel on the Evidence Workshoot	ld be de	esignated with a minus			
(-) S	Symbol on the Evidence Worksneet.					

NEUTRAL (Ø)

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

# **Evidence Worksheet for Primary RESEARCH Article**

<b>Citation:</b> write it in AMA format	Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H., Christle, J., Hastie, T.,
as found in JADA.	Ashley, E. (2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart
	Rate and Energy Expenditure in a Diverse Cohort. Journal of Personalized
	<i>Medicine</i> , 7(2), 3. doi:10.3390/jpm7020003
Study design: Use algorithm –	Validity study
RCT, cohort, etc	
Study Class (A,B,C,D)	С
<b>Research Ouality Rating</b>	POSITIVE (+)
This rating tells if the research	
design is good (+), bad (-) or	
neutral ( $\emptyset$ )	
This is determined by the quality	
criteria list Delete the ratings	
that do not apply (i.e. if positive	
delete minus/negative and	
neutral)	
	Purpose/Population Studied/Practice Studied
	To access the accuracy of access accuracy in the second state of t
Research purpose: What is the	10 assess the accuracy of seven commercially available wrist-worn devices in
research question being	estimating heart rate (HR) and energy expenditure (EE), and to propose a wearable
investigated in the study?)	sensor evaluation framework
Inclusion criteria: requirements	For devices: wrist-worn watch or band, continuous measurement of HR, stated
for study eligibility	battery life greater than 24 hours, commercially available direct to consumer at the
	time of the study, one device per manufacturer
	For participants: healthy adults over 18 years
Exclusion criteria (conditions	For devices: technical problems during pre-testing
that make individual ineligible)	For participants: unclear
Recruitment	Through advertisements within Stanford University and local amateur sports clubs.
	From these interested volunteers, participants were selected to maximize
	demographic diversity as measured by age, height, weight, body mass index (BMI),
	wrist circumference, and fitness level
Blinding used: some of the	NA
persons involved are prevented	
from knowing certain information	
that might lead to conscious or	
unconscious bias on their part,	
invalidating the results	
Description of study protocol	The Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2,
What happened in the study?	PulseOn, and Samsung Gear S2 were evaluated. Participants wore devices while
	being simultaneously assessed with continuous electrocardiography and indirect
	calorimetry while sitting, walking, running, and cycling. Error in HR and EE was
	computed for each subject/device/activity combination.
Intervention: Describe	Devices were tested in two phases; the first phase included the Apple Watch, Basis
interventions, regimens, risk	Peak, Fitbit Surge, and Microsoft Band; the second phase included the MIO Alpha
factors, or procedures studied.	2, PulseOn, and Samsung Gear S2. Participants wore up to four devices and
	simultaneously underwent continuous electrocardiographic monitoring and
	continuous indirect calorimetry using FDA approved equipment. After being fitted
	with all equipment, participants performed the standardized exercise protocol in a
	controlled laboratory setting. The exercise protocol involved five-minute intervals
	of sitting, walking, fast walking, running, fast running, cycling, and intense cycling.
	The running and cycling stages were individualized to each participants' fitness
	level. Data was collected according to manufacturers' instructions or by making

	use of an Application Programming Interface. The last minute of each stage was
	used for data analysis.
<b>Statistical analysis:</b> List tests, significance level set a priori ( $\alpha$ =0.05; include intent to treat analysis if applicable; note if there is Power analysis.	Statistical analyses were performed separately for HR (electrocardiography served as the gold standard) and EE (indirect calorimetry served as the gold standard). Two-way ANOVA with post-hoc Turkey honest significant difference was performed to check for a difference between groups for categorical demographic covariates. A Pearson correlation test was performed between continuous demographic variables and device error. Separate tests were performed for each device, and p-values were adjusted with the Bonferroni correction for multiple testing. Principal component analysis was performed to identify outliers and to cluster devices by error profiles. A singular value decomposition was computed over the activity error rates. Several regression approaches were applied to uncover associations in the dataset, and a Bland-Altman analysis was performed to measure device error relative to the gold standards. An error rate of 5% at a p-value of 0.05 was determined to be within acceptable limits since this approximates a widely accepted standard for statistical and clinical significance.
outcomes were measured; usually baseline and one or more later times	walking, running, and fast running until 25 minutes had passed. Participants then sat for three minutes to rest and recover, followed by five-minute intervals of cycling and intense cycling. Participants were given one minute to sit and recover, concluding the exercise protocol at 39 minutes. HR and EE data were collected continuously throughout the 39-minute intervention. The last minute of each stage was used for data analysis.
<b>Dependent variables:</b> <i>outcomes</i> <i>that are measured or registered;</i> <i>variable whose change or</i> <i>different states the researcher</i> <i>wants to understand, explain, or</i> <i>predict</i>	HR and EE according to the Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, Samsung Gear S2
<b>Independent variables</b> (intervention or procedure; this variable can be manipulated; a variable whose effect upon the dependent variable one is trying to understand)	Participant demographics, such as age, height, weight, BMI, wrist circumference, skin tone, fitness level
Control Variables Examples: 1) multivariate logistic regression controlled for age, BMI, albumin; 2) usual care; 3) isocaloric diet, etc.	HR and EE according to electrocardiography and indirect calorimetry
Initial n (e.g. 731 (298 males, 433 females)) Record number that entered study – not the number screened.	60 participants (29 males and 31 females)
<b>Final n</b> (attrition) number of subjects that completed study	60 participants (29 males and 31 females)
Age usually mean or range	21 to 64 years
Ethnicity (if given)	Unclear, diverse sample
Other relevant demographics:	Skin tone: 1 to 6 (measured by Fitzpatrick skin tone scale)
<i>demographics describe the</i> <i>population (students, athletes, etc)</i>	Fitness level: 31.7 to 66.6 mL/kg/min (measured by VO2 max)
Anthropometrics: e.g. were	Height: 154.4 to 190 cm
groups same or different on	Weight: 47.8 to 130.6 kg

important physical measures	BMI: 17.2 to 39.3 kg/m2
(BMI, fitness level)	Wrist circumference: 13.5 to 21 cm
Location: Where did the study	Stanford University, California
take place? City or country	
Summary of Results: Abstract	All results presented as median and 95% confidence interval. The lowest error in
results including quantitative data	measuring HR was observed for the cycling stage, 1.8% (0.9%–2.7%), while the
and statistics. Include statistical	highest error was observed for the walking stage, 5.5% (3.9%–7.1%). Six devices
significance: P-values, confidence	achieved a median error below 5% in measuring HR for the cycling stage, while the
intervals (CI), relative risk (RR),	Samsung Gear S2 achieved a median error rate of 5.1% (2.3%–7.9%). For the
odds ratios (OR), likelihood ratio,	walking stage, three devices achieved a median error rate below 5%: the Apple
number needed to treat, power	Watch, 2.5% (1.1%–3.9%); the PulseOn, 4.9% (1.4%–8.6%); and the Microsoft
analysis if available.	Band, 5.6% (4.9%–6.3%). The remaining four devices had median error between
	6.5% and 8.8%. Error in estimation of EE was considerably higher than for HR for
	all devices. Median error rates across activities varied from 27.4% (24.0%–30.8%)
	for the Fitbit Surge to 92.6% (87.5%–97.7%) for the PulseOn. For EE, the lowest
	relative error rates across devices were achieved for the walking (31.8% (28.6%–
	35.0%)) and running (31.0% (28.0%–34.0%)) stages, and the highest relative error
	rates across devices were achieved on the sitting stage $(52.4\% (48.9\% - 57.0\%))$ .
	The Apple Watch achieved the lowest overall error in both HR and EE, while the
	Samsung Gear S2 reported the highest. Device error was higher for males, greater
	BMI, and darker skin tone.
	Author's Conclusions
Author conclusion: paraphrase	In a diverse group of individuals, most wrist-worn devices reported HR within $\frac{1}{2}$
that stated by study duthor in	acceptable error range (5%) under controlled laboratory conditions of walking,
body of the report or abstract	reported EE within an accortable arror range under these conditions. A cross
	devices and modes of activities, the Annia Watch had the most favorable error
	profile while the Samsung Gear S2 had the least favorable error profile
	Individuals and practitioners should be aware of the strengths and limitations of
	consumer devices that measure heart rate and estimate energy expenditure. The
	authors encourage transparency from device companies and consistent release of
	validation data to facilitate the integration of such data into clinical care
Reviewer comments: Note	Strengths: zero percent attrition diverse sample of participants with different ages
strengths and limitations of study.	<i>BML</i> and skin tones, the use of numerous consumer and gold standard devices
identify concerns that affect study	standardized exercise protocol and examining several different activity domains
validity and generalizability—	collected by the devices
vour comments should be	
	Limitations: <i>blinding was not utilized</i> , only consumer devices available at the time
<i>italicized</i> )	Limitations: <i>blinding was not utilized</i> , only consumer devices available at the time of this study were tested, and consumer devices were assessed in a controlled
italicized)	Limitations: <i>blinding was not utilized</i> , only consumer devices available at the time of this study were tested, and consumer devices were assessed in a controlled laboratory setting rather than in a free-living environment

Table 3.2.a. Quality Criteria Checklist: Primary Research

DELEVITORE OFFICIANS					
RELEVANCE QUESTIONS	1		N		N
<b>Citation:</b> write it in AMA format as found in JADA (copy and paste from page 1 of worksheet)		Y E	N O	N N	N A
Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H., Christle, J., Hastie, T., Ashley, E.		s		CL	
(2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy				E	
Expenditure in a Diverse Cohort. Journal of Personalized Medicine, 7(2), 3.				A R	
doi:10.3390/jpm7020003					
1. Would implementing the studied intervention or procedure (if found successful) result in	1			Х	
improved outcomes for the patients/clients/population group? (Not Applicable for some					
epidemiological studies)					
2. Did the authors study an outcome (dependent variable) or topic that the patients/clients/	2	Х			
population group would care about?					
3. Is the focus of the intervention or procedure (independent variable) or topic of study a	3		Х		
common issue of concern to dietetics practice?					
4. Is the intervention or procedure feasible (NA for some epidemiological studies)?	4	Х			
If the answers to all of the above relevance questions are "yes", the report is eligible for design	ation w	ith a	plu.	s (+)	on
the Evidence Quality Worksheet, depending on answers to the following validity questions.					
VALIDITY QUESTIONS					
1. Was the research question clearly stated?		Y	1	N U	N
This is usually stated at end of the introduction and just before methods section.		S			А
				L E	
		Х	<u> </u>	A	
1 1 Wee the gravitient intervention (a) or grave advers (in demondent conich la(a)) identified?	1 1	v	<i>r</i>	N	•
This is after called the treatment and emplained in the methode section	1.1	Δ	•		
<i>This is often called the treatment and explained in the methods section.</i>	1.2	v	,		
1.2 was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Δ	•		
These are sometimes called the endpoints; the results section reports the outcomes, but					
Inis information should be in the methods section, too	1.2	v	,		
1.3 were the target population and setting specified?	1.5		•		
The target population is group for whom findings may be applicable; look for this in the					
Introduction and in the methods section		v	1	N U	N
2. Was the <u>selection</u> of study subjects/patients free from blas?		E	(	) N	Α
		5		L	
				E	
		Х	Ľ.	R	
2.1 Were inclusion/exclusion criteria specified (e.g., risk, point in disease progression,	2.1			Σ	Κ
diagnostic or prognosis criteria), and with sufficient detail and without omitting					
criteria critical to the study?					
The authors should give several points about the inclusion/exclusion criteria such as the					
age range of the subjects, disease condition (like hyperlipidemia) required for					
inclusion. Exclusion criteria should be listed, too, although some are					
understood. For example if the ages for inclusion are 18 to 70, the authors will					
probably not specifically note that children and people over age 70 were					
excluded. Most of the time, however, subjects may be excluded for certain					
characteristics such as being pregnant or having some disease (like CHD).					
2.2 Were criteria applied equally to all study groups?	2.2				Х
2.3 Were health, demographics, and other characteristics of subjects described?	2.3	Х	Č.		
There is usually a Table 1 summarizing demographics and characteristics at baseline.					
Groups are <u>not</u> different if the P-Value is $> 0.05$ . If there has been a previous					
paper describing the study population, that paper may be referenced and you		1			
would need to go back to the original publication to see that Table 1.					
2.4 Were the subjects/patients in a representative sample of the relevant population?	2.4	X			
	1	1	1		1

	The abstractor may have to apply a bit of clinical judgment here. Authors try to be brief and may only say that the patients came from the same clinic from people who met the inclusion					
2.1	Criteria.		v	N	U	N
3.	<i>were study groups comparable</i> :		E	0	Ň	A
	There is usually a Table 1 summarizing demographics and characteristics at baseline. Groups are not different if the P-Value is $> 0.05$		3		L	
	Groups are <u>nor</u> all eren if the 1 -v alue is > 0.05.				E A R	Х
	3.1 Was the method of assigning subjects/patients to groups described and unbiased?	3.1				Х
	(Method of randomization identified if RCT)					
	In a strong study, the authors may tell how the subjects were assigned to a group (e.g.					
	randomized block design; or assigned by computer-generated random numbers).					
	Look for instances that show bias; for example I once read a study where patients					
	were randomized to receive liquid energy supplements; however, if someone dialiked their supplement, the supplement allowed to change groups, this is not unbiased.					
	aisliked their supplement, they were allowed to change groups – this is not unbiased!	2.2			v	
	5.2 were distribution of disease status, prognostic factors, and other factors (e.g.,	3.2			Λ	
	should be no significant differences groups at desering? See Table 1 for this - there					
	2.2 Were concurrent controls used? (Concurrent preferred over historical controls)	2.2	v			
	Most RCTs use a concurrent control group Occasionally an intervention study will use a	5.5	Λ			
	prior study as a control group: that is an example of a historical control. That is not					
	as strong a research design as use of concurrent control group A crossover study					
	where the subject acts as his/her own control is use of concurrent control.					
	3.4 If cohort study or cross-sectional study, were groups comparable on important	3.4				Х
	confounding factors and/or were preexisting differences accounted for by using					
	appropriate adjustments in statistical analysis?					
	The groups in a cohort or cross-sectional study should not be different from each other;					
	if they are, a strong study will utilize statistical techniques such as multivariate					
	analyses to remove the variance due to the group differences. Look for this					
	information in the statistics and results sections.					
	3.5 If case control study, were potential confounding factors comparable for cases and	3.5				Х
	controls? If case series or trial with subjects serving as own control, this criterion is					
	not applicable. Criterion may not be applicable in some cross-sectional studies.					
	Subjects are generally matched for age, gender, etc. Look for this in the statistical					
	aescription and results sections.	2.6	v			
	3.6 If diagnostic test, was there an independent blind comparison with an appropriate	3.6	Х			
	Example: comparing body fat analysis method with underwater weighing (gold					
	standard) In studies trying to determine the best equation (like Mifflin-St. Jeor or					
	Harris-Renedict) to predict energy needs a gold standard measure of RFF (Indirect					
	Calorimetry) is used					
4.	Was method of handling withdrawals described?	1	Y	N	U	N
			E S	0	C	А
					L E	
					AR	Х
	4.1 Were follow up methods described and the same for all groups?	4.1				Х
	4.2 Was the number, characteristics of withdrawals (i.e. dropouts, lost to follow up,	4.2				Х
	attrition rate) and/or response rate (cross-sectional studies) described for each group?					
	(Follow up goal for a strong study is 80 %.)					
	This should be found in the results section. If there is attrition $> 20\%$ , it is important to					
1	note that on the worksheet (as a note in the results section or in the reviewer					
	comments at the very bottom)		4	<u> </u>		
	4.3 Were all enrolled subjects/patients (in the original sample) accounted for?	4.3	Х	1		
	This information is often presented in a figure with # recruited, # enrolled (this is initial			1		
	N), # remaining at end of study period (final N). Sometimes the reasons that subjects			1		
	withdrew or were dropped is given in the figure or in the text (results section).			1		

-						
	4.4 Were reasons for withdrawals similar across groups?	4.4				Х
	If there is a large attrition from one group and not others, you would want to look for a					
	reason why; the answer to this question would then be no.	4.5	v			
	4.5 If diagnostic test, was decision to perform reference test not dependent on results of test under study?	4.5	A			
	The test under study should be compared to reference test all the time. An example of this					
	might be using a DEXA machine to measure percent body fat only if a subject's					1
	BMI was $>$ 35 but bioimpedance analyzer indicated body fat $<$ 30%.		N/	N		N
5.	Was <u>blinding</u> used to prevent introduction of bias?		Y E	N O	U N	N A
			s		C L A R X	
	5.1 In intervention study, were subjects, clinicians/practitioners and investigators blinded	5.1				Х
	to treatment group, <u>as appropriate</u> ?					ı
	The key term is as appropriate. For example, in the Lim et al 2008 study, the investigators					
	studied the effect of MNT on lipid levels in hypercholesterolemic patients. It was					i i
	an RCT, but obviously, the subjects and practitioners knew who was getting MNT					ı
	and who was not. Therefore, you would not answer question 5.1 NO. It was					
	appropriate for the dictitians and patients to know they were receiving MNT.	5.2			37	
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is measured	5.2			Х	
	using an objective test, such as a lab value, this criterion is assumed to be met.)					ı
	Answer yes, if a lab lest was used to measure an outcome. A method of blinding a diel					i i
	collecting the data)					
	5.3 In cohort study or cross-sectional study, were measurements of outcomes and risk	5.3				Х
	factors blinded?					
	Answer yes, if a lab test was used to measure an outcome. A method of blinding a diet					
	study is to have separate people analyzing the data (not the same ones who were collecting the data).					1
	5.4 In case control study, was case definition explicit and case ascertainment not influenced by exposure status?	5.4				Х
	Establish who is a case and who is a control at the beginning of the study.					
	5.5 In diagnostic study, were test results blinded to patient history and other test results?	5.5			Х	
6.	Were intervention/therapeutic regimens/exposure factor or procedure and any comparison(s) described in detail? Were intervening factors described?		Y E S	N O	U N C L	N A
			x		E A	
	6.1 In RCT or other intervention trial, were protocols described for all regimens studied?	6.1			R	Х
	6.2 In observational study, were interventions, study settings, and clinicians/provider	6.2	X			
	described?					
	6.3 Was the intensity and duration of the intervention or exposure factor sufficient to produce a meaningful effect?	6.3	Х			
	Use clinical judgment (e.g. 12 weeks is long enough for a dietary intervention to make a					ı
	difference in lab values for cholesterol; however, 12 days would not be long					
	enough)					
	6.4 Was the amount of exposure and, if relevant, subject/patient compliance measured?	6.4	Х			
	How long did the treatment last? Did the patient follow directions?					
	6.5 Were co-interventions (e.g., ancillary treatments other therapies) described?	6.5				Х
	(e.g. were patients on lipid-lowering meds at the same time as the diet therapy)	<u> </u>		$\square$		
	6.6 Were extra or unplanned treatments described?	6.6				Х
	The text may not describe any unplanned treatments. If yes, it would likely be in the					I
	asscussion section. It is likely there were no unplanned treatments, so a "no"					
	unswer is not a problem overall.					

	6.7 Was the information for 6.4, 6.5, 6.6 and 6.7 assessed the same way for all groups? <i>For a study to be valid and unbiased, it is important that this be yes.</i>	6.7				Х
	6.8 In diagnostic study, were details of test administration and replication sufficient? <i>Usually answer n/a for diet study.</i>	6.8	Х			
7.	Were <u>outcomes</u> clearly defined and the measurements valid and reliable?		Y E S	N O	U N C L E A	N A
	<ul> <li>7.1 Were primary and secondary endpoints described and relevant to the question?</li> <li>Primary endpoint -main result measured at the end of a study to see if the treatment worked. The primary endpoint is decided at the beginning of the study.</li> <li>Secondary endpoint - not as important as the main results; not usually analyzed if the primary endpoint is not statistically significant.</li> </ul>	7.1	X		R	
	<ul> <li>7.2 Were nutrition measures appropriate to question and outcomes of concern?</li> <li><i>Clinical judgment required: weight loss, changes in energy intake are relevant to MNT;</i> Sometimes there are no nutrition measures and you should answer N/A.</li> </ul>	7.2				Х
	7.3 Was the period of follow-up long enough for important outcome(s) to occur? <i>Clinical judgment required: was there enough time?</i>	7.3				Х
	<ul> <li>7.4 Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?</li> <li>Check that surveys were validated.</li> </ul>	7.4	X			
	7.5 Was the measurement of effect at an appropriate level of precision? Precision is reproducibility or repeatability	7.5	X			
	<ul> <li>7.6 Were other factors accounted for (measured) that could affect outcomes?</li> <li>Other factors are sometimes covered in the discussion of the strengths/limitations of the study.</li> </ul>	7.6	X			
	7.7 Were the measurements conducted consistently across groups?	7.7				Х
8.	Was the <u>statistical analysis</u> appropriate for the study design and type of outcome indicators?		Y E S X	N O	U N C L E A R	N A
	8.1 Were statistical analyses adequately described and the results reported appropriately? <i>There should be a discussion of the statistics in the methods section.</i>	8.1	Х			
	<ul> <li>8.2 Were correct statistical tests used and assumptions of test not violated?</li> <li>You will get better at this the more papers you abstract.EAL abstractors are expected to have some statistical and research training (minimum of master's degree).</li> </ul>	8.2	X			
	8.3 Were statistics reported with levels of significance and/or confidence intervals? ( <i>P-value</i> ) and/or confidence intervals (mean $\pm$ CI)	8.3	Х			
	<ul> <li>8.4 Was "intent to treat" analysis of outcomes done (and as appropriate, was there an analysis of outcomes for those maximally exposed or a dose-response analysis)?</li> <li>Intent to treat– analysis is based on the original treatment intent, not the treatment ultimately administered (i.e. does not matter if treatment was for 2, 6, 8 or all the weeks in the study). The analyses are done using all the subjects in the study, not just the ones who completed it. This is done in order to avoid effects of dropout that can be a threat to randomization. <i>Intent-to-treat analysis of outcomes applies to any intervention study</i>. If intent to treat analysis was done, it will be mentioned in the statistical section. If all subjects who began the trial completed it, intent-to-treat analysis was done.</li> </ul>	8.4				X
	<ul> <li>8.5 Were adequate adjustments made for effects of confounding factors that might have affected the outcomes (e.g. multivariate analyses)?</li> <li>Multivariate analyses are used to adjust or control for other variables (age, sex, smoking, etc). Assumes data is valid and reduces a larger number of variables to a smaller number. Just answer yes or no that multivariate analyses were used.</li> </ul>	8.5	X			

8.6 Was clinical significance as well as statistical significance reported? <i>Example: Lim, et al 2008 reported that after 12 weeks of MNT, total cholesterol was</i> <i>reduced from 229.2±158 to 181.3±16.3 (P&lt;0.001); This includes: statistical</i> <i>significance (P-value) and clinical significance (compare to standard of &lt; 200</i> <i>mg/do for normal cholesterol). A problem can occur when only statistical</i> <i>significance is reported. Reducing cholesterol from 300 to 250 might be statistically</i> <i>significant, but clinically it is still abnormal.</i>	8.6	X			
<ul> <li>8.7 If negative findings, was a power calculation reported to address type 2 error?</li> <li>Type II (β error is a false negative that happens when the investigators fail to reject the <u>null hypothesis</u> when the null hypothesis is false. Look for the authors to say something like "a sample size of n=xx is needed to provide 80% power."</li> </ul>	8.7				Х
9. Are <u>conclusions</u> supported by results with biases and limitations taken into consideration?		Y E S	N O	U N C L E	N A
		Х		A R	
9.1 Is there a discussion of findings? Answer yes or no.	9.1	Х			
9.2 Are biases and study limitations identified and discussed? This will be in the discussion of finding section that follows the results	9.2	Х			
<b>10.</b> Is bias due to study's <u>funding or sponsorship</u> unlikely? Be careful here – if <u>bias</u> is <u>unlikely</u> , answer YES.		Y E S X	N O	U N C L E A R	N A
<ul> <li>10.1 Were sources of funding and investigators' affiliations described?</li> <li>Look just under the abstract, or</li> <li>The funding may be acknowledged at the end of the paper</li> <li>Just because the work was funded by industry does not mean the study was biased.</li> </ul>	10.1	X			
<ul><li>10.2 Was there no apparent conflict of interest?</li><li>If an investigator is testing a piece of equipment, process or drug that s/he developed, it could be a conflict of interest.</li></ul>	10.2	X			
SYMBOL					
MINUS/NEGATIVE (-) If most (six or more) of the answers to the above validity questions are "no," the report should minus (-) symbol on the Evidence Quality Worksheet. NEUTRAL (a)	be design	nated	with	n a	
INEU INAL (0) If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is except	tionally s	trong	the	,	

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Quality Worksheet.

#### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" including criteria 2, 3, 6, and 7 and at least one additional "yes", (the report should be designated with a plus symbol (+) on the Evidence Quality Worksheet.

Citation	Shcherbina, A., Mattsson, C., Waggott, D., Salisbury, H., Christle, J., Hastie, T., Ashley, E. (2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. Journal of Personalized Medicine, 7(2), 3. doi:10.3390/jpm7020003
Study Design	Validity study
Class	C
Quality Rating	$\square$ + (Positive) $\square$ - (Negative) $\square$ $\otimes$ (Neutral)
Research Purpose	To assess the accuracy of seven commercially available wrist-worn devices in estimating heart rate (HR) and energy expenditure (EE), and to propose a wearable sensor evaluation framework
Inclusion Criteria	For devices: wrist-worn watch or band, continuous measurement of HR, stated battery life greater than 24 hours, commercially available direct to consumer at the time of the study, one device per manufacturer For participants: healthy adults over 18 years
Exclusion Criteria	For devices: technical problems during pre-testing For participants: unclear
Description of Study Protocol	Recruitment: Through advertisements within Stanford University and local amateur sports clubs. From these interested volunteers, participants were selected to maximize demographic diversity as measured by age, height, weight, body mass index (BMI), wrist circumference, and fitness level Design: The Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2 were evaluated. Participants wore devices while being simultaneously assessed with continuous electrocardiography and indirect calorimetry while sitting, walking, running, and cycling. Error in HR and EE was computed for each subject/device/activity combination. Blinding used (if applicable): NA Intervention (if applicable): Devices were tested in two phases; the first phase included the Apple Watch, Basis Peak, Fitbit Surge, and Microsoft Band; the second phase included the MIO Alpha 2, PulseOn, and Samsung Gear S2. Participants wore up to four devices and simultaneously underwent continuous electrocardiographic monitoring

	and continuous indirect calorimetry using FDA approved equipment.
	After being fitted with all equipment, participants performed the
	standardized exercise protocol in a controlled laboratory setting. The
	exercise protocol involved five-minute intervals of sitting, walking, fast
	walking, running, fast running, cycling, and intense cycling. The running
	and cycling stages were individualized to each participants' fitness level.
	Data was collected according to manufacturers' instructions or by making
	use of an Application Programming Interface. The last minute of each
	stage was used for data analysis.
	Statistical Analysis: Statistical analyses were performed separately for
	HR (electrocardiography served as the gold standard) and EE (indirect
	calorimetry served as the gold standard). Two-way ANOVA with post-
	hoc Turkey honest significant difference was performed to check for a
	difference between groups for categorical demographic covariates. A
	Pearson correlation test was performed between continuous demographic
	variables and device error. Separate tests were performed for each device,
	and p-values were adjusted with the Bonferroni correction for multiple
	testing. Principal component analysis was performed to identify outliers
	and to cluster devices by error profiles. A singular value decomposition
	was computed over the activity error rates. Several regression approaches
	were applied to uncover associations in the dataset, and a Bland-Altman
	analysis was performed to measure device error relative to the gold
	standards. An error rate of 5% at a p-value of 0.05 was determined to be
	within acceptable limits since this approximates a widely accepted
	standard for statistical and clinical significance.
	Timing of Measurements: The exercise protocol involved five-minute
Data Collection	intervals of sitting, walking, fast walking, running, and fast running until
Summary	25 minutes had passed. Participants then sat for three minutes to rest and
	recover, followed by five-minute intervals of cycling and intense cycling.
	Participants were given one minute to sit and recover, concluding the

	exercise protocol at 39 minutes. HR and EE data were collected
	continuously throughout the 39-minute intervention. The last minute of
	each stage was used for data analysis.
	Dependent Variables: HR and EE according to the Apple Watch, Basis
	Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, Samsung
	Gear S2
	Independent Variables: Participant demographics, such as age, height,
	weight, BMI, wrist circumference, skin tone, fitness level
	Control Variables: HR and EE according to electrocardiography and
	indirect calorimetry
	Initial: 60 (29 Males 31 Females)
	Attrition (final N): 60
	Age: 21-64 years
	Ethnicity: Unclear, diverse sample
Description of	Other relevant demographics: Skin tone: 1 to 6 (measured by Fitzpatrick
Actual Data Sample	skin tone scale), fitness level: 31.7 to 66.6 mL/kg/min (measured by VO2
	max)
	Anthropometrics: Height: 154.4 to 190 cm, weight: 47.8 to 130.6 kg,
	BMI: 17.2 to 39.3 kg/m2, wrist circumference: 13.5 to 21 cm
	Location: Stanford University, California
	Key Findings: All results presented as median and 95% confidence
	interval. The lowest error in measuring HR was observed for the cycling
	stage, 1.8% (0.9%–2.7%), while the highest error was observed for the
	walking stage, 5.5% (3.9%–7.1%). Six devices achieved a median error
	below 5% in measuring HR for the cycling stage, while the Samsung Gear
Summary of	S2 achieved a median error rate of 5.1% (2.3%–7.9%). For the walking
Results	stage, three devices achieved a median error rate below 5%: the Apple
	Watch, 2.5% (1.1%–3.9%); the PulseOn, 4.9% (1.4%–8.6%); and the
	Microsoft Band, 5.6% (4.9%–6.3%). The remaining four devices had
	median error between 6.5% and 8.8%. Error in estimation of EE was
	considerably higher than for HR for all devices. Median error rates across

	activities varied from 27.4% (24.0%–30.8%) for the Fitbit Surge to 92.6%					
	(87.5%–97.7%) for the PulseOn. For EE, the lowest relative error rates					
	across devices were achieved for the walking (31.8% (28.6%-35.0%))					
	and running (31.0% (28.0%-34.0%)) stages, and the highest relative error					
	rates across devices were achieved on the sitting stage (52.4% (48.9%-					
	57.0%)). The Apple Watch achieved the lowest overall error in both HR					
	and EE, while the Samsung Gear S2 reported the highest. Device error					
	was higher for males, greater BMI, and darker skin tone.					
	Other Findings:					
	In a diverse group of individuals, most wrist-worn devices reported HR					
	within acceptable error range (5%) under controlled laboratory conditions					
	of walking, running, and cycling at low and high intensities. None of the					
	wrist-worn devices reported EE within an acceptable error range under					
	these conditions. Across devices and modes of activities, the Apple					
Author	Watch had the most favorable error profile while the Samsung Gear S2					
Conclusion	had the least favorable error profile. Individuals and practitioners should					
	be aware of the strengths and limitations of consumer devices that					
	measure heart rate and estimate energy expenditure. The authors					
	encourage transparency from device companies and consistent release of					
	validation data to facilitate the integration of such data into clinical care.					
	Strengths: zero percent attrition, diverse sample of participants with					
	different ages, BMI, and skin tones, the use of numerous consumer and					
	gold standard devices, standardized exercise protocol, and examining					
Reviewer	several different activity domains collected by the devices					
Comments	Limitations: blinding was not utilized, only consumer devices available at					
	the time of this study were tested, and consumer devices were assessed in					
	a controlled laboratory setting rather than in a free-living environment					
Funding Source	Unclear, the authors declare no conflict of interest					

Quality Criteria Checklist: Primary ResearchSymbols UsedExplanation

+	<b>Positive</b> – Indicates that the report has clearly addressed issues of inclusion/exclusion, bias, generalizability, and data collection and analysis
	Negative – Indicates that these issues have not been adequately addressed.
0	<i>Neutral</i> – indicates that the report is neither exceptionally strong nor exceptionally week

Select a rating from the drop-down menu ↓

Re	levance Questions		
1.	Would implementing the studied intervention or procedure (if found successful) result in improved outcomes for the patients/clients/population group? (NA for some Epi studies)	1	Unclear
2.	Did the authors study an outcome (dependent variable) or topic that the patients/clients/population group would care about?	2	Yes
3.	Is the focus of the intervention or procedure (independent variable) or topic of study a common issue of concern to dietetics practice?	3	No
4.	Is the intervention or procedure feasible? (NA for some epidemiological studies)	4	Yes

If the answers to all of the above relevance questions are "Yes," the report is eligible for designation with a plus (+) on the Evidence Quality Worksheet, depending on answers to the following validity questions.

Va	alidity	Questions		
1.	Was th	e <u>research question</u> clearly stated?	1	Yes
	1.1.	Was the specific intervention(s) or procedure (independent variable(s))	1.1	Yes
	1.2.	Was the outcome(s) (dependent variable(s)) clearly indicated?	1.2	Yes
	1.3.	Were the target population and setting specified?	1.3	Yes
2.	Was th	ne <u>selection</u> of study subjects/patients free from bias?	2	Yes
	2.1.	progression, diagnostic or prognosis criteria), and with sufficient detail and	2.1	Unclear
	2.2.	without omitting criteria critical to the study? Were criteria applied equally to all study groups?	2.2	N/A
	2.3.	Were health, demographics, and other characteristics of subjects described?	2.3	Yes
	2.4.	population?	2.4	Yes
3.	Were <u>9</u> 3.1.	Study groups comparable? Was the method of assigning subjects/patients to groups described and	3	N/A
	3.2.	unbiased? (Method of randomization identified if RCT) Were distribution of disease status, prognostic factors, and other factors (e.g., demographics) similar across study groups at baseline?	3.1	N/A
	3.3.	Were concurrent controls used? (Concurrent preferred over historical controls.)	3.2	Unclear
	3.4.	If cohort study or cross-sectional study, were groups comparable on important confounding factors and/or were preexisting differences accounted for by using	3.3	Yes
	3.5.	appropriate adjustments in statistical analysis? If case control study, were potential confounding factors comparable for cases and controls? (If case series or trial with subjects serving as own control, this	3.4	N/A
			3.5	N/A

	criterion is not applicable. Criterion may not be applicable in some cross-		
	sectional studies.)	3.6	Ves
	3.6. If diagnostic test, was there an independent blind comparison with an	5.0	105
	appropriate reference standard (e.g., "gold standard")?		
4.	Was method of handling <u>withdrawals</u> described?	4	N/A
	4.1. Were follow up methods described and the same for all groups?	4 1	
	4.2. Was the number, characteristics of withdrawals (i.e., dropouts, lost to follo	W 4.1	IN/A
	up, attrition rate) and/or response rate (cross-sectional studies) described	for 4.2	N/A
	4.3 Were all enrolled subjects/nations (in the original sample) accounted for?	4.3	Yes
	4.4. Were reasons for withdrawals similar across groups		
	4.5. If diagnostic test, was decision to perform reference test not dependent of	า	IN/A
	results of test under study?	4.5	Yes
5.	Was <u>blinding</u> used to prevent introduction of bias?	5	Unclear
	5.1. In intervention study, were subjects, clinicians/practitioners, and investiga	tors	
	5.2 Were data collectors blinded for outcomes assessment? (If outcome is	5.1	N/A
	measured using an objective test, such as a lab value, this criterion is assu	med	
	to be met.)	5.2	Unclear
	5.3. In cohort study or cross-sectional study, were measurements of outcomes risk factors blinded?	and 5.3	N/A
	5.4. In case control study, was case definition explicit and case ascertainment r influenced by exposure status?	ot 5.4	N/A
	5.5. In diagnostic study, were test results blinded to patient history and other t	est 5.5	Unclear
	results?	5.5	Ulicical
6.	Were <u>intervention</u> /therapeutic regimens/exposure factor or procedure and any	6	Yes
	comparison(s) described in detail? Were <u>intervening factors</u> described?		
	studied?	5 0.1	IN/A
	6.2. In observational study, were interventions, study settings, and	6.2	Yes
	clinicians/provider described?	6.3	Yes
	6.3. Was the intensity and duration of the intervention or exposure factor suffi	cient	1 00 XX
	6.4 Was the amount of exposure and if relevant subject/natient compliance	6.4	Yes
	measured?	6.5	N/A
	6.5. Were co-interventions (e.g., ancillary treatments, other therapies) describe	ed? 6.6	N/A
	6.6. Were extra or unplanned treatments described?		
	6.8. In diagnostic study, were details of test administration and replication	6.7	N/A
	sufficient?	6.8	Yes
7.	Were <u>outcomes</u> clearly defined and the <u>measurements valid and reliable</u> ?	7	Yes
	7.1. Were primary and secondary endpoints described and relevant to the question?	7.1	Yes
	7.2. Were nutrition measures appropriate to question and outcomes of concer	n? 7.2	N/A
	7.3. Was the period of follow-up long enough for important outcome(s) to occu	ır? 7.3	N/A
	7.4. Were the observations and measurements based on standard, valid, and reliable data collection instruments/tests/procedures?	7.4	Yes
	7.5. Was the measurement of effect at an appropriate level of precision?	7.5	Yes
	7.6. Were other factors accounted for (measured) that could affect outcomes?		37
		7.6	l Yes
	7.7. Were the measurements conducted consistently across groups?	7.6	Yes N/A

8.	Was t	ne <u>statistical analysis</u> appropriate for the study design and type of outcome	8	Ves
	indica	tors?		105
	8.1.	Were statistical analyses adequately described the results reported	8.1	Yes
		appropriately?	0 1	Var
	8.2.	Were correct statistical tests used and assumptions of test not violated?	0.2	res
	8.3.	Were statistics reported with levels of significance and/or confidence intervals?	8.3	Yes
	0.4.	an analysis of outcomes for those maximally exposed or a dose-response analysis)?	8.4	N/A
	8.5.	Were adequate adjustments made for effects of confounding factors that	8.5	Yes
		might have affected the outcomes (e.g., multivariate analyses)?	8.6	Yes
	8.6.	Was clinical significance as well as statistical significance reported?		105
	8.7.	If negative findings, was a power calculation reported to address type 2 error?	8.7	N/A
9.	Are <u>co</u>	nclusions supported by results with biases and limitations taken into	9	Yes
	consid	eration?	9.1	Yes
	9.1.	Is there a discussion of findings?	0.0	V
	9.2.	Are biases and study limitations identified and discussed?	9.2	Yes
10.	Is bias	due to study's <u>funding or sponsorship</u> unlikely?	10	Yes
	10.1	. Were sources of funding and investigators' affiliations described?	10.1	Yes
	10.2	. Was there no apparent conflict of interest?	10.2	Yes
MI If m (-)	NUS/N 10st (six symbol UTRA	EGATIVE (-) or more) of the answers to the above validity questions are "No," the report shoul on the Evidence Worksheet. L (Ø)	d be de	esignated with a minus
If +		and the solidity endeavies an extension 2.2.6 and 7 do not in direct the state of a state in the	: 1	1

If the answers to validity criteria questions 2, 3, 6, and 7 do not indicate that the study is exceptionally strong, the report should be designated with a neutral ( $\emptyset$ ) symbol on the Evidence Worksheet.

### PLUS/POSITIVE (+)

If most of the answers to the above validity questions are "Yes" (including criteria 2, 3, 6, 7 and at least one additional "Yes"), the report should be designated with a plus symbol (+) on the Evidence Worksheet.

Author,	Study Purpose	Study	Intervention	Outcomes	Limitations
Year,		Population			
Study					
Design,					
Class,					
Rating	Ta agagagin a	15 (7 malas	Doutioinouto	The Eithit Change 2	Small somels
Author: Banadatto at	10 assess in a	15 (7 males	rade a	avhibited a mean bias	size participant
al	research	females)	stationary bike	of 5.0  hpm (05%)	size, participant
ai. Vear: 2018	environment	healthy	for 10 minutes	CI) The limits of	not discussed
Study	the accuracy	Caucasian	with the stated	agreement $(I \circ A)$	hlinding was not
Design <sup>.</sup>	and precision of	adult	goal to raise	between the Fithit	utilized possible
Validity	the Fitbit	participants	their HR as	Charge 2 and	unstable
study	Charge 2 for	Age: 25 to 36	much as	ProComp Infiniti	positioning of the
Class: C	measuring heart	vears	possible.	T7500M were wide.	Fitbit Charge 2.
Rating: +	rate (HR) with	Weight: 56 to	Participants'	The upper LoA was	lacking a defined
	respect to a	82 kg	HR was	+16.8 bpm, whereas	activity pattern for
	gold standard	Height: 155	simultaneously	the lower LoA was -	the participants to
	electrocardiogra	to 185 cm	recorded from	28.5 bpm. The	simulate low,
	ph	BMI: 20 to	the Fitbit	intraclass correlation	medium, and
		25 kg/m2	Charge 2 and	coefficients (ICC)	intensive exercise,
			ProComp	Charge 2 and	and lacking a
			Infiniti 1/500M	Charge 2 and	variety of
				T7500M was 0.21	participants
				(95% CI)	
Author:	To determine	40 (20 males	Participants	The Fithit Surge had	Participant
Cadmus-	the accuracy of	and 20	wore two	the best LoA $(-5.1 \text{ to})$	recruitment.
Bertram et	the heart rate	females)	activity trackers	4.5 beats/min) while	demographics,
al.	measured by	healthy adult	on each wrist	the Basis Peak had	and location of
Year: 2017	four	participants	and were	the worst LoA (-17.1	study were not
Study	commercial,	Age: 30 to 65	connected to an	to 22.6 beats/min)	discussed, and
Design:	light-emitting	years	electrocardiogra	while resting. When	blinding was not
Validity	diode-	Mean BMI:	ph. Participants	participants	utilized
study	dependent,	25.1 kg/m2	sat and rested	exercised, the LoA	
Class: C	wrist-worn		for 10 minutes,	were relatively poor	
Kating: Ø	(Basis Beak		on a treadmill at	trackers (Mio Euse	
	Eithit Charge		65% of their	-225  to  260	
	Fitbit Surge		maximum heart	beats/min: Basis	
	Mio Fuse)		rate for 10	Peak, -27.1 to 29.2	
	, í		minutes while	beats/min; Fitbit	
			their heart rates	Surge, -34.8 to 39.0	
			were measured.	beats/min; and Fitbit	
				Charge, -41.0 to 36.0	
1		51 (0 1		beats/min)	
Author:	To evaluate the	51 (0  males)	Participants	After the 16-week	Small sample
Caamus-	reasibility and	and 51	were	Fithit group increased	size, short
al	integrating a	nostmenonau	aither a Eithit or	FILDIL GLOUP INCREASED $MVPA hy 62\pm108$	of generalizability
Year: 2015	Fithit tracker	sal	nedometer-	$\min/week (n < 0.01)$	since narticinants
Study	and website into	overweight or	based	MVPA in 10-min	were all
Design:	a physical	obese women	intervention	bouts by 38±83	postmenopausal.

Randomized controlled trial Class: A Rating: +	activity intervention for postmenopausal , overweight or obese women	Age: 53 to 67 years BMI: 25.7 to 32.7 kg/m2	group to determine whether the Fitbit One increased physical activity more than the pedometer	min/week (p=0.008), and steps by 789±1,979 (p=0.01), compared to non- significant increases in the pedometer group (between- group p-values: 0.11, 0.28 and 0.30).	overweight or obese women and there were several confounders such as the goal setting process, four- week follow-up call, and Fitbit website
Author: Chum et al. Year: 2017 Study Design: Randomized controlled trial Class: A Rating: ø	To understand patients' perceived benefit from the Fitbit One and explore themes associated with patient experiences. To compare perceived benefit, patient factors, Fitbit usage, and Beck's Depression Inventory (BDI) scores.	36 (18 males and 18 females) participants with major depressive disorder (MDD) Mean age: 53 years Mean BDI score: 36.27 (scale of 0- 63) Mean BMI: 30.16 kg/m2	Fitbit One activity trackers were provided to all patients allocated to the behavioral activation (BA) group at the beginning of the BRAVE study. Patients were encouraged to carry the Fitbit One at all times throughout 18 weeks to track their physical activity. Interviews were conducted with 36 patients who completed the BRAVE study.	23 patients found the Fitbit One to be helpful for their physical activity. Themes of positive experiences included self-awareness, peer motivation, and goal- setting opportunities. Themes of negative experiences included inconvenience, inaccuracies, discouragement, and disinterest. There was a significant relationship between total Fitbit One usage and perceived benefit. The mean number of weeks of Fitbit One use for those who found the Fitbit helpful was 18.57 and 12.27 weeks for those who did not (p<0.001).	Small sample size compared to the number of participants who completed the BRAVE study, difficulty interpreting exploratory findings given that the study was underpowered to test effectiveness, and there were several confounders such as combined BA intervention, study setting, and participants' restricted use of Fitbit
Author: Cook et al. Year: 2017 Study Design: Validity study Class: C Rating: +	To evaluate the utility of the Fitbit Flex (FBF) to estimate sleep in adult patients with MDD relative to gold standard polysomnograp hy (PSG) and validated actigraphy (Actiwatch-2; AW-2).	21 (4 males and 17 females) unmedicated participants with MDD Mean age: $26.5 \pm 4.6$ years Mean BDI-II score: $22.9 \pm 6.8$	Patients wore the FBF and AW-2 during in-laboratory PSG. The following sleep variables were calculated: total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE).	Compared to PSG, the FBF significantly overestimated TST (mean difference of 46.0 min, p<0.0001) and SE (mean difference of 8.1%, p<0.0001), while significantly underestimating WASO (mean difference of -44.0 min, p<0.0001). SOL assessed by FBF and PSG were quite similar (mean difference of -2.0 min, p=0.72). The FBF showed a high sensitivity (0.98 $\pm$	Participants were young to middle aged and predominantly female. Study specifically examined outpatients with MDD, limiting the generalizability of findings. Results cannot be extended to other fitness trackers, or more current generations of the same model as these devices may have different

				0.02) and accuracy (0.88 $\pm$ 0.05), with low specificity (0.35 $\pm$ 0.13).	performance characteristics.
Author: Ferguson et al. Year: 2015 Study Design: Validity study Class: C Rating: +	To assess the concurrent validity of a selection of consumer-level accelerometer- based activity monitors (Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike Fuelband, Striiv Smart Pedometer, and Withings Pulse) compared to two research- grade accelerometers (BodyMedia SenseWear, and ActiGraph GT3X+) in free-living conditions	21 (10 males and 11 females) healthy adult participants Age: 20 to 59 years Male BMI: $27.3 \pm 3.2$ kg/m2 Female BMI: $25.5 \pm 5.2$ kg/m2	Participants wore seven activity monitors and two research- grade accelerometers simultaneously for 48-hours. Participants went about their daily life in free-living conditions. Four physical activity parameters were measured, including step count, moderate to vigorous physical activity (MVPA), sleep, and total daily energy expenditure (TDEE)	All activity monitors measured steps, and correlations with reference devices were very strong (r = 0.94-0.99). Five activity monitors measured MVPA, and correlations ranged from weak to strong (r = $0.52$ - 0.91). Four activity monitors measured sleep, and all correlated strongly with the reference device (r = $0.82$ - 0.92). Five activity monitors measured TDEE, and correlations were moderate to strong (r = $0.74-0.81$ ). The Fitbit One, Fitbit Zip, and Withings Pulse were the strongest performers.	Participant recruitment was not discussed, blinding was not utilized, and validity may vary if activity monitors are worn in locations other than the hip or wrist
Author: Gomersall et al. Year: 2016 Study Design: Validity study Class: C Rating: Ø	To compare Fitbit One and Jawbone UP estimates of steps, MVPA, and sedentary behavior with data from the ActiGraph GT3X+ accelerometer in a free-living context	32 (3 males and 26 females) healthy adult participants Mean age: 39.6, SD: 11.0 years Mean BMI: 25.9, SD: 5.0 kg/m2	On two occasions for seven days each, participants wore an ActiGraph GT3X+ on their hip and either a hip-worn Fitbit One or wrist- worn Jawbone UP. Daily estimates of steps, MVPA, and longest sedentary time were measured.	Correlations for steps and MVPA were strong for both devices, although higher for the Fitbit One ( $r$ =.85 for steps and $\rho$ =.80 for MVPA) than for Jawbone UP ( $r$ =.75 for steps and $\rho$ =.75 for MVPA). The correlation between the Jawbone UP longest idle time and ActiGraph longest sedentary bout was poor ( $\rho$ =.19). Agreement between the Fitbit One and ActiGraph for the classification of active versus inactive was substantial ( $P$ <.001), while	Predominantly female, healthy, middle-aged sample which limits the generalizability of the findings, and the study could not control for wear time of the activity trackers

				agreement between the Jawbone UP and ActiGraph was	
Author: Gualtieri et al. Year: 2016 Study Design: Non- randomized crossover trial Class: C Rating: ø	To determine (1) if participants would accept and use activity trackers to increase their physical activity; (2) if there were barriers to use besides cost and training; (3) if activity trackers would educate participants on their activity levels and support behavior change; and (4) if clinical outcomes would show improvements in participants' health	10 (2 males and 8 females) adult participants with one chronic medical condition Age: 39 to 77 years	Patients were given Withings Pulse wearable activity trackers in the physician-led wellness group and were interviewed two to four weeks after the 12- week study. Themes about participants attitudes and behavior changes were analyzed along with participants' clinical data over the course of the study.	moderate ( $P$ <.001). Improvements were seen in clinical outcomes, physical activity behaviors, and attitudes towards the Withings Pulse. Participants lost an average of 0.5 pounds per week, with a mean total weight loss of 5.97 pounds ( $P$ =.004). Other clinical outcomes included a 9.2% decrease in LDL levels ( $P$ =.038). Changes in blood pressure were non- significant. All participants reported an increase in well- being, health education, physical activity, and confidence in their ability to lead more active lives	Small sample size, predominantly female sample, study design lacked a control group, blinding was not utilized, and the results cannot separate the impact of the wellness group education and support from that of the activity tracker use
Author: Maher et al. Year: 2017 Study Design: Cross- sectional study Class: D Rating: +	To explore users' experience of activity trackers, including the perceived usefulness of devices for tracking and modifying lifestyle behaviors (physical activity, diet, and sleep), ease of use, patterns of use, and barriers to use	237 (69 males and 168 females) adult participants who were current or former activity tracker users Age: 18 to 74 years	A cross- sectional online survey was developed to address the research objectives and was administered via Survey Monkey to adults who were current or former activity tracker users	The most commonly used brand of activity tracker was Fitbit (67.5%), followed by Garmin (16.5%), Apple (3.4%), Jawbone (2.5%), Samsung (1.7%), Polar (1.3%), and other (7.1%). Participants agreed that various features on their trackers were useful, including: steps (95%), active minutes (76%), sleep (66%), heart rate (63%), stairs climbed (58%), and energy burned (57%), while fewer agreed that the food intake feature was useful (36%). Overall, 94% of current users and	Relatively high dropout rate, predominantly female sample, study design increased the risk of recall bias, and difficulty knowing how generalizable the results are

				65% of former users agreed that they had	
				had a positive	
				experience using	
Author	To examine	34(13  males)	Participants	Darticipants achieved	Small cample
Naslund et	whether	and 21	wore Fithit Zins	an average of 4453 5	size
al.	average daily	females) non-	most of the	steps each day, with	predominantly
Year: 2016	step count	Hispanic	days they were	average daily step	female sample,
Study	measured using	white, obese,	enrolled in the	counts ranging from	lacking racial or
Design:	Fitbit Zip	adult	6-month group	1037.6 to 11,366.3	ethnic diversity,
Non-	devices was	participants	behavioral	steps. There was a	relatively high
crossover	associated with	services for	program	association between	blinding was not
trial	weight loss and	schizophrenia	Participants'	participants' average	utilized, analyses
Class: C	improved	spectrum	weight and	daily step count and	were based on
Rating: +	fitness among	disorder,	change in	weight loss. For	participants who
	individuals with	MDD, or	fitness was	every 1000 step	completed the
	illness enrolled	disorder	haseline and 6	experienced a	study, results
	in a 6-month	Mean age:	months. Daily	decrease in weight of	the impact of
	lifestyle	50.2 years,	step count data	1.78 pounds ( $p =$	group education
	program	SD = 11.0	was extracted	0.0314). The	and support from
		Mean weight:	from	relationship between	the use of Fitbit
		231.9 lbs Mean BMI	Fitbite Zipe	average daily step	Zips, and findings
		38.5  kg/m2	Thous Zips	fitness was not	representative of
		Fitness:		significant (increase	individuals with
		1303.8 feet in		of 18.79 feet on the	serious mental
		6-Minute		6-MWT (p = 0.176)).	illness not
		Walk Test (6-			currently
Author:	To compare the	40(19  males)	Participants	Mean error analyses	Inclusion and
Rosenberger	output from	and 21	wore nine	for the devices	exclusion criteria
et al.	commercially	females)	devices for 24-	ranged from 8.1% for	and participant
Year: 2016	available	adult	hours:	GT3X+ to 16.9% for	demographics
Study	wearable	participants	Actigraph	GENEactiv when	were not
Validity	current	Age. 21 to 70 years	activPAL Fitbit	duration: 9 5% for	blinding was not
study	standards for	yeurs	One,	LUMOback to 65.8%	utilized, standards
Class: C	objective		GENEactiv,	for GENEactiv when	were based on
Rating: ø	measurement of		Jawbone Up,	measuring SED;	common field-
	sleep, sedentary		LUMOback, Nike Euelband	19.7% for GENEactive to 28.0%	based measures
	(SED) light-		Omron	for Fitbit when	standards used in
	intensity		pedometer, and	measuring LPA;	the laboratory,
	physical		Z-Machine.	51.8% from Jawbone	placement of
	activity (LPA),		Comparisons	to 92.0% from	activity monitors
	(MVPA), and		(to standards)	rueidand when measuring MVPA.	can affect how
	living		total sleen time	and 14.1% from	match up to
	environment.		(Z-machine),	GT3X+ to 29.9%	standards, and the
			time spent in	from Fuelband when	results cannot be
			· · · · · · ·		
			SED	measuring total steps	extended to other

			(GT3x+),	only one comparison	generations of the
			MVPA duration	device, the	same model as
			(GI3x+), and	LUMOback was	these devices may
			total steps per	significantly	have different
			day (Omron).	equivalent to	hardware and
				standards for SED	software updates.
				(90% CI).	
Author:	To assess the	60 (29 males	Devices were	The lowest error in	Blinding was not
Shcherbina	accuracy of	and 31	evaluated in	measuring HR was	utilized, only
et al.	seven	females)	two phases.	observed for the	consumer devices
Year: 2017	commercially	diverse,	Participants	cycling stage, 1.8%	available at the
Study	available wrist-	healthy adult	wore up to four	(0.9%-2.7%, 95%	time of th study
Design:	worn devices	participants	devices while	CI), while the highest	were tested, and
Validity	(Apple Watch,	Age: 21 to 64	being	error was observed	consumer devices
study	Basis Peak,	years	simultaneously	for the walking stage,	were assessed in a
Class: C	Fitbit Surge,	Height: 154.4	assessed with	5.5% (3.9%–7.1%,	controlled
Rating: +	Microsoft Band,	to 190 cm	continuous	95% CI). Error in	laboratory setting
	Mio Alpha 2,	Weight: 47.8	electrocardiogra	estimation of EE was	rather than in a
	PulseOn, and	to 130.6 kg	phy and indirect	considerably higher	free-living
	Samsung Gear	BMI: 17.2 to	calorimetry	than for HR for all	environment
	S2) in	39.3 kg/m2	while sitting,	devices. Median	
	estimating heart	Skin tone: 1	walking,	error rates across	
	rate (HR) and	to 6	running, and	activities varied from	
	energy	(measured by	cycling. Error	27.4% (24.0%-	
	expenditure	Fitzpatrick	in HR and EE	30.8%, 95% CI) for	
	(EE), and to	skin tone	was computed	the Fitbit Surge to	
	propose a	scale)	for each	92.6% (87.5%-	
	wearable sensor	Fitness level:	device/activity	97.7%, 95% CI) for	
	evaluation	31.7 to 66.6	combination.	the PulseOn. The	
	framework	mL/kg/min		Apple Watch	
		(measured by		achieved the lowest	
		VO2 max)		overall error, while	
				the Samsung Gear S2	
				reported the highest.	

Draduat	Apple Watch Sories	Fithit Charge	Fithit Vorso	Cormin	Comin	
Frounce	Apple watch series	riton Charge	FILDIL VEISA	Vivosmart 4	Vivosport	
	T ttps://www.bhphotovideo.com/c/product/ 134390. REG/apple_mtug2l1 a_watch_series_4_gp <u>s.html</u>	Attps://www.wallmart.com/jp/ Fithit-Charge-3-Advanced- Hart-Rate-Fritness- Tracker/654994366	https://www.amazon.com/Fit bit/yersas-fama-familiami included/dp:B07B48SQGTp1 b_1	https://www.amazon.com/Gar min-v%C3%ADvosmart- ActivityFitness- Midnight/dp/B07GM7WHBG	https://www.clevertraining.com/gar min-vivosport-gps-activity-tracker	
Release Date	September 2018	October 2018	April 2018	September 2018	August 2017	
Price	\$399	\$149.95	\$199.95	\$129.99	\$169.99	
Wear Site	Wrist	Wrist	Wrist	Wrist	Wrist	
Compatibility	iOS	Android, iOS, Windows	Android, iOS, Windows	Android, iOS	Android, iOS	
Display	OLED	OLED	Color LCD	OLED	Color LCD	
Battery	18 hours	7 days	4+ days	7 days	7 days	
Water	Yes	Yes	Yes	Yes	Yes	
Resistant						
Functions						
Steps	Yes	Yes	Yes	Yes	Yes	
Distance	Yes	Yes	Yes	Yes	Yes	
Elevation/	Yes	Yes	Yes	Yes	Yes	
Stairs Heart Data	Vaa	Vaa	Vaa	Vaa	Vac	
Colorios	T es Vec	Vec	Vec	I es Vec	I es Vec	
Burned	1 05	1 65	1 05	1 05	1 05	
Active Time	Yes	Yes	Yes	Yes	Yes	
Sleep Time	Yes	Yes	Yes	Yes	Yes	
Sleep Quality	Yes	Yes	Yes	Yes	Yes	
Other	Apps, music, GPS, notifications, goal setting, exercise modes, coaching, activity sharing, guided breathing	Apps, notifications, goal setting, exercise modes, activity sharing, guided breathing	Apps, music, notifications, goal setting, exercise modes, coaching, activity sharing, guided breathing	Apps, music, notifications, goal setting, VO2 max, pulse ox, exercise modes, activity sharing, guided breathing	Apps, music, GPS, notifications, goal setting, VO2 max, exercise modes, activity sharing	

## APPENDIX C: COMPARISON REPORT

Product	Misfit Shine	Moov Now	Samsung Gear	Withings Steel	Xiaomi Mi
	2		Fit2	HR Sport	Band 3
	https://www.amazon.com/ Misfit.Shine-Fitness- Tracker-	https://www.amazon.com/Fitm ess-Tracker-Audio-Couch- Moor/dp/B01CX26E0	https://www.samsung.com/global/ga laxy/gear-fit2/	https://www.smartwatchspex.com/with ings-sted-hr-sport-specifications/	https://www.amazon.co.uk/Wristh and-Fitness-incoming-waterproof- forcegate/db/07D1001
Release Date	November	November	June 2016	September 2018	May 2018
	2015	2015		~ · · · · · · · · · · · · · · · · · · ·	
Price	\$79.99	\$49.99	\$179.99	\$199.95	\$29.99
Wear Site	Wrist	Wrist	Wrist	Wrist	Wrist
Compatibility	Android, iOS, Windows	Android, iOS	Android, iOS	Android, iOS	Android
Display	12 color LED lights	None	AMOLED	Analog dial, subdial, OLED	OLED
Battery	6 months	6 months	3+ days	25 days	20 days
Water	Yes	Yes	Yes	Yes	Yes
Resistant					
Functions			Ι	Γ	I
Steps	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes
Elevation/ Stairs	No	Yes	Yes	No	No
Heart Rate	No	Yes	Yes	Yes	Yes
Calories	Yes	Yes	Yes	Yes	Yes
Burned					
Active Time	Yes	Yes	Yes	Yes	Yes
Sleep Time	Yes	Yes	Yes	Yes	Yes
Sleep	Yes	Yes	Yes	Yes	Yes
Quality					
Other	Music,	Exercise	Music, GPS,	Notifications, VO2	Notifications,
	notifications,	modes,	notifications, goal	max, exercise	goal setting,
	goal setting,	coaching	setting, exercise	modes	exercise modes,
	activity		modes, activity		activity sharing
	sharing		sharing		