HHS Public Access

Author manuscript

Transl J Am Coll Sports Med. Author manuscript; available in PMC 2023 October 12.

Published in final edited form as:

Transl J Am Coll Sports Med. 2022 ; 7(4): 1–7. doi:10.1249/tjx.0000000000000215.

Systematic Review of Fitbit Charge 2 Validation Studies for Exercise Tracking

Crista Irwin[1], **Rebecca Gary**[1]

[1]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA

Abstract

Context—There are research-grade devices that have been validated to measure either heart rate (HR) by electrocardiography (ECG) with a Polar chest strap, or step count with ACTiGraph accelerometer. However, wearable activity trackers that measure HR and steps concurrently have been tested against research-grade accelerometers and HR monitors with conflicting results. This review examines validation studies of the Fitbit Charge 2 (FBC2) for accuracy in measuring HR and step count and evaluates the device's reliability for use by researchers and clinicians.

Design—This registered review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The robvis (risk-of-bias visualization) tool was used to assess the strength of each considered article.

Eligibility Criteria—Eligible articles published between 2018 and 2019 were identified using PubMed, CINHAL, Embase, Cochran, and World of Science databases and hand-searches. All articles were HR and/or step count validation studies for the FBC2 in adult ambulatory populations.

Study Selection—Eight articles were examined in accordance with the eligibility criteria alignment and agreement among the authors and research librarian.

Main Outcome Measures—Concordance correlation coefficients (CCC) were used to measure agreement between the tracker and criterion devices. Mean absolute percent error (MAPE) was used to average the individual absolute percent errors.

Results—Studies that measured CCC found agreement between the FBC2 and criterion devices ranged between 26% and 92% for HR monitoring, decreasing in accuracy as exercise intensity increased. Inversely, CCC increased from 38% to 99% for step count when exercise intensity increased. HR error between MAPE was 9.21% to 68% and showed more error as exercise intensity increased. Step measurement error MAPE was 12% for healthy persons aged 24–72 years but was reported at 46% in an older population with heart failure.

Corresponding author: Crista Irwin, MS, BSN, RN, cevan01@emory.edu, Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Rd., Atlanta, GA 30322.

Conflicts of Interest and Sources of Funding

The authors of this project have no conflicts of interest to declare. This project was partially supported by the T32 training grant in Interventions to Improve Outcomes in Chronic Conditions (T32NR012715–06, Dunbar S, Song M (co-PIs)), and by the Healing Hearts and Mending Minds "Fitbrain" study (R-01 NR014963, Gary R and Waldrop D (co-PIs)).

Conclusions—Relative agreement with criterion and low-to-moderate MAPE were consistent in most studies reviewed and support validation of the FBC2 to accurately measure HR at low or moderate exercise intensities. However, more investigation controlling testing and measurement congruency is needed to validate step capabilities. The literature supports the validity of the FBC2 to accurately monitor HR, but for step count is inconclusive so the device may not be suitable for recommended use in all populations.

Keywords

Fitbit Charge 2; exercise tracker; heart rate; step count

Introduction

Background

Current universal guidelines recommend at least 150 minutes of weekly moderate intensity exercise of between 3–6 metabolic equivalent of task (MET) for persons across the health strata (1). Moderate intensity exercise can be achieved by brisk walking at a pace of 100 steps per minute or measured at 60–70% of heart rate (HR) maximum (2). Persons with chronic conditions who routinely participate in physical activity at moderate intensity for a minimum of 150 minutes per week exhibit significant improvements in cardiovascular function and reduction in inflammation (1,3–12). Incorporating exercise prescriptions into healthcare protocols show major benefits among a variety of health conditions including cardiovascular disease, depression, and HIV, as well as for preventative medicine (13). Providers need cost-effective, reliable wearable devices that accurately measure physical activity to assess and provide feedback to patients regarding their physical activity and to fully utilize the benefits of an exercise prescription.

Current research-grade devices have been validated to measure either HR using electrocardiography (ECG) (14), step count using an ACTigraph accelerometer (15), or sensing HR pulse using a Polar chest strap (16), but to date no wearable device has been validated to measure both HR and steps concurrently. Exercise equipment manufacturers continuously develop multifunction activity trackers worn as watches that purport to accurately capture steps and HR; however, these wearable activity trackers have been tested against research-grade accelerometers and heart rate monitors with variable results. The Fitbit Charge 2 (FBC2) (Fitbit Inc., San Francisco, CA, USA) is a low-cost, wearable device that tracks steps and HR and has been studied in a variety of populations in home-based and healthcare settings. The purpose of this review is to examine validation studies of the FBC2 for accurately measuring HR and step count and to evaluate the device's reliability to determine whether the device can be recommended by healthcare providers for use by patients.

Features of the Fitbit Charge 2

Fitbit trackers use microelectronic triaxial accelerometer and proprietary algorithms to measure step gait and distance and continuous light-emitting diode (LED) lighting to measure pulse continually. They are multifunctional, wrist-worn devices that not only measure steps and HR but include a multitude of user-friendly features. The device must

be wirelessly connected via Bluetooth to a network-connected mobile phone. Through this connectivity, the device may receive text and call notifications. The device may be connected to the owner's contact list to develop community support networks for exercise motivation with special permission. The device software often sends supportive messages to encourage movement throughout the day or once the owner achieves personal activity goals set by him/herself. The software package also includes workout videos that can be accessed on the mobile phone application. The device also functions as a watch and has timer, mileage, relaxation, and stopwatch features.

Limitations of the Fitbit Charge 2

Because the device requires smart-phone and internet access and has an average price of $150, the FBC2 may be a difficult option for low-income populations. Many basic and low-functioning mobile phones lack the capability to support the Fitbit application. The watch is rechargeable and includes a charging cord, which is easily misplaced and/or broken, and needs 6–7 hours charging time which lasts for approximately 3–4 days. Consumers report frequent watchband and equipment failure after 12–18 months of use. Additionally, the manufacturer does not report whether the updated models, which are released about every 18 months, have been altered significantly and, therefore, require updated testing and validation.

Methods

Design

This review was registered on the International Prospective Register of Systematic Reviews (PROSPERO) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used in conducting the review and reporting the appropriate articles (17). Figure 1 illustrates the PRISMA flow diagram of included articles. Eight articles on the validation of FBC2 in adult ambulatory populations that were published between 2018 and 2019 were examined for this review. Seven of the articles were randomized control trials and one was a test of a single participant who wore multiple trackers. Articles were excluded if they did not assess the FBC2 model, did not assess HR or step counts, or if the population sample was under 18 years of age. The relatively small article publication time range is due in part to the speed at which commercially available wearable exercise tracker technology changes.

Strategy

The methods for this review included a search of PubMed, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochran, Embase, World of Science, hand searches, and assistance by Emory University's Health Sciences Librarian. Key words and combinations of the words used in the search were "Fitbit Charge 2," "exercise tracker," "activity tracker," "activity monitor," "heart rate," "steps," and "validation." Article inclusion criteria were HR and/or steps validation studies for the FBC2 in adult ambulatory populations.

Data Extraction

Each validation article included in this review was required to assess HR and step count accuracy. Articles were also included if they evaluated intra-reliability testing between the FBC2 trackers used in the testing as well as inter-reliability among other trackers assessed. Criterion measures for HR were based on the research-validated ECG and the Polar chest strap for ambulatory activities (14,15). Step count data was compared to the validated ACTiGraph wGT3X-BT (16).

To assess the tracker's accuracy, seven of the eight studies in this review explained the differences in the data collected by the FBC2 using the following measures (Table1): mean error, as in the difference between the criterion measure and the consumer device; mean absolute error, as in the average absolute distance between the data from the consumer and the criterion devices; mean percent error or relative error rate, as in the difference between the criterion measure and the consumer device, represented by a percentage; and mean absolute percent error (MAPE), as in the average of the individual absolute percent errors. MAPE analyzes individual overestimation and underestimation values taken by the device and, therefore, may be a more appropriate representation of the activity monitors when comparing studies.

Concordance correlation coefficient (CCC) 95% confidence intervals (CI) were used to describe strength of agreement between the devices in four studies (18–20). Interclass correlation coefficient (ICC) is used when comparing multiple devices to each other and to the criterion. Additionally, because precise data congruency collected between consumer devices is unlikely, Bland-Altman (BA) analysis was used by four studies to evaluate proximity to data measured by criterion devices (18,20,23,24).

Another article was included in this validation review because the authors tested the FBC2 tracker's accuracy for measuring cardiorespiratory fitness compared to maximal oxygen uptake (VO_{2max}) (25). Cardiorespiratory fitness (CRF), defined as the circulatory and respiratory systems' transport and utilization of oxygen to the skeletal muscles, is typically measured by maximal graded exercise testing on a treadmill and measured in units of $mL \cdot kg^{-1} \cdot min^{-1}$. Under strict laboratory protocols by a trained exercise physiologist using precise equipment, the gold standard for evaluating CRF levels is by VO_{2max} (26). Researchers have found that low levels of CRF measured by VO_{2max} treadmill testing have been associated with cardiovascular disease risk (25,27,28). The FBC2 purports to evaluate CRF using proprietary algorithms which include an individual user's age, weight, height, resting HR, and peak HR.

Articles were evaluated by two independent researchers and inclusion agreement was discussed in detail. Validation articles included in this review are listed and described in Tables 1 and 2.

Risk of bias assessment

Each study was evaluated for risk of bias (Table 2). Criteria used for the assessment included randomization bias, recruitment bias, protocol deviation bias including criterion tool bias, missing outcome bias, and reporting bias. The selected studies were consistently evaluated

as high quality based on these criteria with few concerns for risk of bias. Most of the concerns were regarding racially homogenous or small samples. One study included a single participant and most of the studies included majority white persons and healthy populations. The risk-of-bias visualization tool (robvis; Bristol, AC, UK) was used to create the risk of bias assessment (Fig. 2) (29).

Results

Accuracy testing

Of the eight studies found, five reported MAPE values of the FBC2 against the criterion for each study which are considered acceptable at <10%. The MAPE values for the studies evaluating HR accuracy include 9.21% (18), 10.79% (19), and 69% (24). Mean absolute percent error values investigating step count acuity were 12.36% (22), 46% for participants with heart failure (HF), and 12% among healthy controls (21). Another study measured the HR difference between the FBC2 and the criterion using relative error rate (RER) (light activity, 5.36%, to moderate activity, 9.3%) (20).

Intra-reliability testing

A 24-hour evaluation study found 91% CCC (95% CI 0.896, 0.914) agreement of the criterion with the FBC2 in a single participant (male, 29 years) (18). Another study cited 92% CCC (95% CI 0.92, 0.93) agreement of the criterion with the FBC2 in their randomized controlled trial including 20 participants (mean age 27.5 yr, standard deviation (SD) 6 yr; 55% female) while walking on a treadmill (19). Additionally, researchers collected data from 30 participants (mean age 23.5 yr, SD 3 yr; 50% female) and found that, as exercise intensity increased, agreement decreased (20). At very low HR, intensity ranging between 55 and 90 beats per minute (bpm), CCC agreement between the criterion and the FBC2 was 89% (95% CI 0.79, 0.95) (20). When HR ranged between 90 and 120 bpm, CCC agreement was moderate at 55%, (96% CI 0.28, 0.74) and CCC was poor at 26% (95% CI 0.01, 0.46) when HR ranged between 110 and 150 bpm (20). In contrast, step counting criterion reliability increased with treadmill speeds in a 3-day field study of 15 participants (mean age 65.5 yr, SD 12.6 yr; 40% female) with heart failure compared to 14 (mean age 43 yr, SD 18.9 yr; 64% female) healthy controls (21). CCC agreement of the step criterion with the FBC2 was 38% (95% CI 0.00, 0.67) at 2.4 km/h (slow walk), 82% (95% CI 0.68, 0.97) at 3.0 km/h (moderate walk), and 99% (95% CI 0.98, 1.0) at 3.6 km/h (brisk walk) (21).

Inter-reliability testing

In a 24-hour study of 20 (mean age 70.2 yr, SD 2.9 yr; 55% female) older healthy adults in Ireland, results showed >0.89 ICC strength of agreement between devices for step count evaluation (22). In addition, a study from Korea including 51 participants (mean age 44.4 yr, SD 16.6 yr; 53% male; 100% Asian) who were undergoing electrophysiological study and ablation to treat paroxysmal tachycardia or supraventricular tachycardia found >0.98 ICC strength of agreement between the FBC2 and the ECG for HR monitoring (23). Recorded baseline HR monitoring with the FBC2 was within ±5 beats per minute of the criterion ECG at 95% accuracy (23). However, device agreement of the FBC2 and criterion results by

Pearson correlations assessed in other studies in the review were incongruent measuring 0.23 (poor) and 0.94 (equivalent), respectively (19,24).

Cardiorespiratory fitness assessment

Researchers compared the VO_{2max} values obtained from the standard treadmill tests with the CRF values estimated by the FBC2 (25). In a sample of 65 healthy adults aged 18–45 yr (55% female), Bland-Altman analyses showed that the FBC2 CRF had a positive bias of 1.59 $mL \cdot kg^{-1} \cdot min^{-1}$ when compared to the treadmill testing at 15 s and a positive bias of 0.30 $mL \cdot kg^{-1} \cdot min^{-1}$ at 60 s with MAPE values <10% for each comparison (26).

Discussion

Heart rate validation

Four studies in this review assessed the FBC2 for HR accuracy validation. All except one study included healthy participants, aged 21–73 yr, and generally reported more accuracy at lower-intensity activity levels (18–23). Heart rate MAPE values while walking at low-to-moderate-intensity levels, at 9.21%, 10.79%, and 69%, reveal a wide interval of error results, with two of the three being similar (18,19,24). The RER reported by one group of researchers supports validation with their statistically moderate error rate at low-to-moderate walking intensity (light activity, 5.36%, to moderate activity, 9.3%) (20). Another study used pacing cycle length (PCL) data obtained during scheduled electrophysiological studies to evaluate the HR accuracy of the FBC2 (23). At 100 bpm, the FBC2 measured within ±5 bpm when compared to the ECG criterion at a rate of 93% accuracy with atrial pacing and 80% accuracy with ventricular pacing (23). However, the FBC2 device became significantly less accurate at higher bpm (23). HR and steps inherently fluctuate with intensity. These results are similar to the other studies reviewed.

Step count validation

In the Irish study of older adults, the FBC2 overestimated step count (MAPE 12.36%, approaching the acceptable range of <10%), but had vastly different results than the study comparing the older HF subjects (MAPE 46%) to younger healthy controls (MAPE 12%) (21,22). The explanation for why the MAPE values of the two healthy populations in the studies were similar while the MAPE values among the HF participants showed much higher error rates is unclear. However, alterations in gait and slower walking speed among the HF patients likely challenge the FBC2 to track steps reliably and may be a concern when using this device to track steps in populations with ambulation limitations or considerable exercise intolerance due to symptom severity.

FBC2 as HR monitor

Reliability results as determined by criterion agreement with the FBC2 reported in this review were markedly varied. Scores <0.50 indicate poor reliability, 0.50 to 0.75 moderate reliability, and >0.75 good reliability (30). Nelson et al. (18) and Reddy et al. (19) reported high CCC scores of >90%. However, Thomson et al. (20) showed decreasing reliability from 56% (moderate) to 26% (poor) as HR intensity increased. Pearson coefficient results from two studies revealed the widest reliability agreement strength discrepancy from

0.23 (weak) and 0.94 (equivalent) (19,24). Finally, Bland Altman analysis plots revealed HR underestimation measured by the FBC2 compared to criterion at all intensity levels (18,20,24). The differences of the results may be caused by erratic arm movements or misplacement of the tracker bands as the participants move and perspiration. These varied results make it difficult to reach a definitive conclusion regarding reliability across intensity levels, but support reliability at low-to-moderate exercise dose levels.

Fitbit Charge 2 as step counter

Reliability of the FBC2 is in agreement with the step criterion, which is the opposite of the HR results. CCC agreement increased from 38% at lower speeds to 99% at a brisk walk (21). ICC results of >0.89 supports evidence for high agreement strength between the FBC2 step counter and actigraphy (22).

Cardiovascular fitness validation

Researchers reported that the FBC2 could be validated to evaluate CRF in relatively young, healthy persons, especially those with a high level of fitness (26). Nearly 92% of the total participants in this study were classified as having high CRF (26). Although the study found CRF agreement between the FBC2 and the Balke treadmill test among users with lower fitness levels, the low numbers in the “good” or “poor” fitness-leveled groups sampled in the study do not provide sufficient evidence of variation to determine validity nor is the sample representative of the general population who are typically less engaged in CRF activities. Validation studies are needed in populations with chronic conditions or who have ambulation challenges to further evaluate the CRF feature in the FBC2.

Study limitations

Validation consensus of the FBC2 is limited due to the studies’ small sample sizes (n=1 to n=60) and non-standardized activity settings with some conducted in laboratories on treadmills and others in free-living conditions. Most of the HR examinations were only conducted using young and healthy subjects and may not be generalizable to populations who have chronic conditions or to older adults with other physical limitations. The study that compared HF subjects to healthy controls included far different age demographics (21). In addition, the review is limited by the small number of relevant studies available within a short time span which is due in part to the development speed of new technology. The Fitbit company released the FBC2 in 2016 and the FBC3 became available in 2018. The cost and research effort needed to perpetually study and validate new technology limits the viability of commercial wearable devices such as the FBC2 for research and use in primary care. Developers of commercial devices would benefit monetarily from strategic collaborations with healthcare researchers in producing devices that are technologically consistent and reliable. There is great potential for wide use of more accessible and affordable devices by healthcare providers worldwide.

Conclusion

Although the FBC2 has been validated for moderate HR and step count accuracy in some studies, more investigation controlling testing and measurement congruency is needed to

validate both HR and step capabilities. The literature supports the validity of the FBC2 to accurately monitor HR at low-to-moderate exercise intensities, but validation for step count is inconclusive and may not be suitable for recommended use by populations with gait speed or ambulation challenges.

Acknowledgements

The authors would like to acknowledge Sharon Leslie, the librarian for the Nell Hodgson Woodruff School of Nursing at Emory University, for her guidance in devising a strategic plan for this review. We also appreciate the support of Jeannine Cimiotti for her editorial and educational contributions. The results of the study do not constitute endorsement by ACSM.

References

1. Office of Disease Prevention and Health Promotion. Physical activity guidelines [Internet]. Washington (DC): US Department of Health and Human Services; 2021 [cited 2021 Dec 31]. Available from: https://health.gov/paguidelines/.
2. Marshall SJ, Levy SS, Tudor-Locke CE, et al. Translating physical activity recommendations into a pedometer-based step goal 3000 steps in 30 minutes. Am J Prev Med 2009;36(5):410–5. [PubMed: 19362695]
3. Sallam N, Laher I. Exercise Modulates Oxidative Stress and Inflammation in Aging and Cardiovascular Diseases. Oxid Med Cell Longev. 2016;2016:7239639.
4. Hotting K, Roder B. Beneficial effects of physical exercise on neuroplasticity and cognition. Neurosci Biobehav Rev. 2013;37(9 Pt B):2243–57. [PubMed: 23623982]
5. Nwaka II, Onwunzo CN, Ezeakunne U, et al. Effect of aerobic exercises and resistance exercises on inflammatory biomarkers and fatigue in people living with HIV infection undergoing highly active antiretroviral therapy: A systematic review. Indian J Physiother Occup Ther. 2019;13(4):180–5.
6. Pedro RE, Candido N, Guariglia DA, et al. Exercise improves cytokine profile in HIV-infected people: A randomized clinical trial. Cytokine. 2017;99:18–23. [PubMed: 28683357]
7. Roos R, Myezwa H, van Aswegen H, Musenge E. Effects of an education and home-based pedometer walking program on ischemic heart disease risk factors in people infected with HIV: a randomized trial. J Acquir Immune Defic Syndr. 2014;67(3):268–76. [PubMed: 25140908]
8. Zheng G, Qiu P, Xia R, et al. Effect of aerobic exercise on inflammatory markers in healthy middle-aged and older adults: A systematic review and meta-analysis of randomized controlled trials. Front Aging Neurosci. 2019;11:98. [PubMed: 31080412]
9. Owusu C, Nock NL, Hergenroeder P, et al. IMPROVE, a community-based exercise intervention versus support group to improve functional and health outcomes among older African American and non-Hispanic White breast cancer survivors from diverse socioeconomic backgrounds: Rationale, design and methods. Contemp Clin Trials. 2020;92:106001.
10. Vingren JL, Curtis JH, Levitt DE, et al. Adding resistance training to the standard care for inpatient substance abuse treatment in men with human immunodeficiency virus improves skeletal muscle health without altering cytokine concentrations. J Strength Cond Res. 2018; 32(1):76–82. [PubMed: 29257793]
11. Fazeli PL, Woods SP, Heaton RK, et al. An active lifestyle is associated with better neurocognitive functioning in adults living with HIV infection. J Neurovirol. 2014;20(3): 233–42. [PubMed: 24554483]
12. Ibeneme SC, Omeje C, Myezwa H, et al. Effects of physical exercises on inflammatory biomarkers and cardiopulmonary function in patients living with HIV: a systematic review with meta-analysis. BMC Infect Dis. 2019;19(1):359. [PubMed: 31035959]
13. Luan X, Tian X, Zhang H, et al. Exercise as a prescription for patients with various with various diseases. J Sport Health Sci. 2019;8(5) 422–41. [PubMed: 31534817]
14. Willemsen GH, De Geus EJ, Klaver CH, et al. Ambulatory monitoring of the impedance cardiogram. Psychophysiology. 1996;33(2):184–93. [PubMed: 8851246]

15. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. 2011;14(5):411–6. [PubMed: 21616714]
16. Gillinov S, Etiwy M, Wang R, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. Med Sci Sports Exerc. 2017;49(8):1697–703. [PubMed: 28709155]
17. Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6(7):e1000097.
18. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. JMIR Mhealth Uhealth. 2019;7(3):e10828.
19. Reddy RK, Pooni R, Zaharieva DP, et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR Mhealth Uhealth. 2018;6(12):e10338. [PubMed: 30530451]
20. Thomson EA, Nuss K, Comstock A, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J sports sci 2019;37(12):1411–9. [PubMed: 30657025]
21. Vetrovsky T, Siranec M, Marencakova J, et al. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. PLoS ONE. 2019;14(9):e0222569.
22. Tedesco S, Sica M, Ancillao A, et al. Environment in an older cohort. JMIR Mhealth Uhealth. 2019;7(6):e13084. [PubMed: 31219048]
23. Hwang J, Kim J, Choi KJ, et al. Assessing accuracy of wrist-worn wearable devices in measurement of paroxysmal supraventricular tachycardia heart rate. Korean Circ J 2019;49(5):437–45. [PubMed: 30808083]
24. O'Driscoll R, Turicchi J, Hopkins M, et al. The validity of two widely used commercial and research-grade activity monitors, during resting, household, and activity behaviours. Health Technol. 2020;10:637–48.
25. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation. 2016;134(24):e653–99. [PubMed: 27881567]
26. Klepin K, Wing D, Higgins M, et al. Validity of cardiorespiratory fitness measured with Fitbit compared to VO2max. Med Sci Sports Exerc. 2019;51(11):2251–6. [PubMed: 31107835]
27. Gupta S, Rohatgi A, Ayers CR, et al. Cardiorespiratory fitness and classification of risk of cardiovascular disease mortality. Circulation. 2011;123(13):1377–83. [PubMed: 21422392]
28. Mandsager K, Harb S, Cremer P, et al. Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. JAMA Netw Open. 2018;1(6):e183605. [PubMed: 30646252]
29. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods. 2021;12(1):55–61. [PubMed: 32336025]
30. Portney LG. Foundations of clinical research: applications to evidenced-based practice. Philadelphia (PA): F.A. Davis; 2020. 696 p.

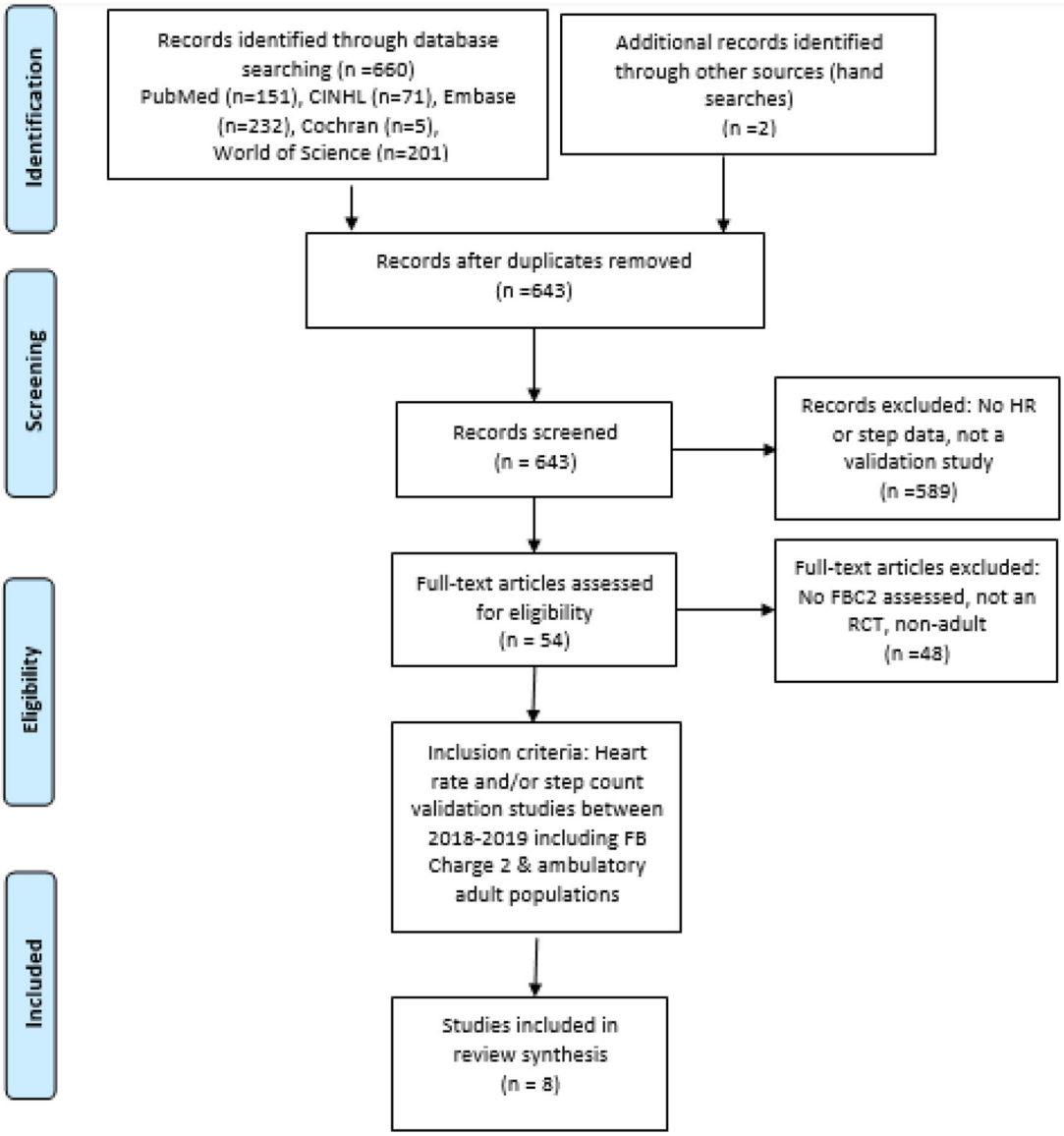

Fig. 1.
PRISMA flow diagram for systematic review of Fitbit Charge 2 validation studies

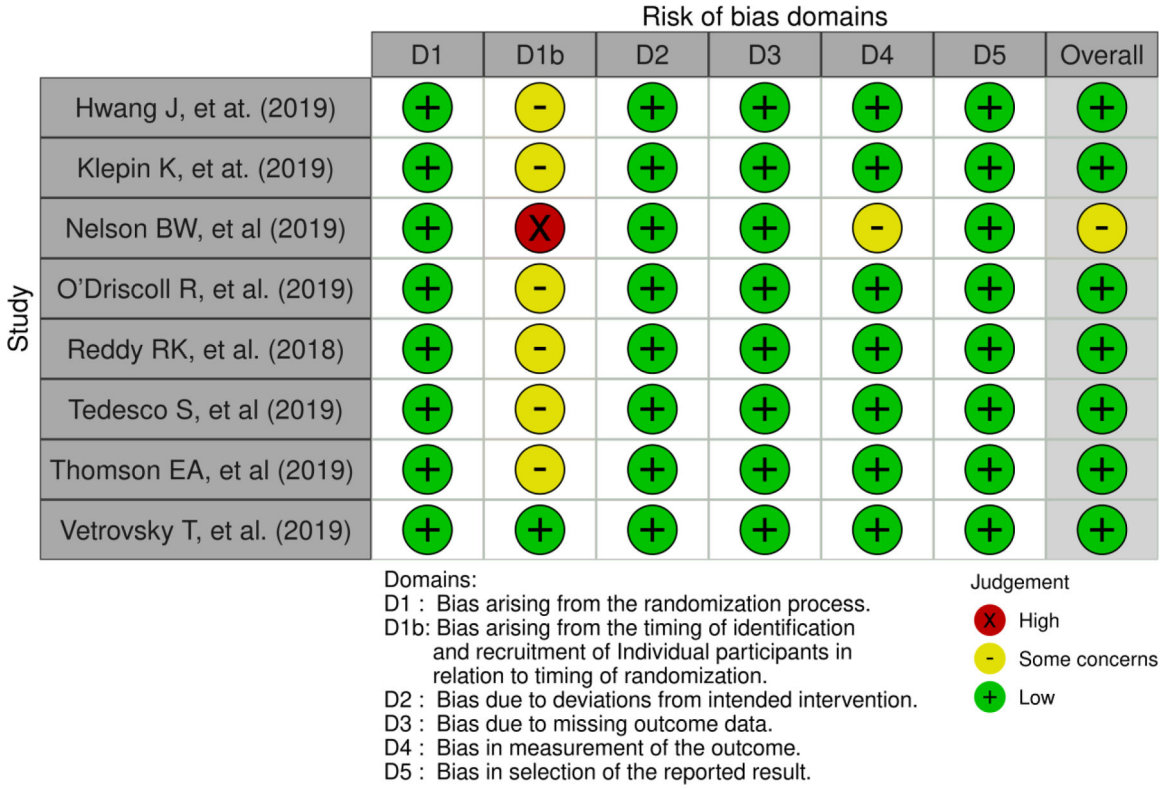

Fig. 2.
Risk of bias assessment

Table 1:

Summary of measurements

Study author/year	Criterion measure	HR	Steps	CCC	ICC	PC	BA	ME	MAE	MPE/RER	MAPE	Risk of Bias
Hwang J, et at. (2019)	ECG	X			X		X					LOW
Klepin K, et at. (2019)	VO_{2max}	X			X						X	SOME CONCERNS
Nelson BW, et al (2019)	ECG	X		X			X		X		X	SOME CONCERNS
O'Driscoll R, et al. (2019)	Polar chest strap	X				X	X		X		X	LOW
Reddy RK, et al. (2018)	Polar chest strap	X		X		X		X		X	X	LOW
Tedesco S, et al (2019)	ActiGraph wGT3X-BT and New-Lifestyles NL2000i		X		X			X		X	X	LOW
Thomson EA, et al (2019)	ECG	X		X			X			X		LOW
Vetrovsky T, et al. (2019)	ActiGraph wGT3X-BT		X	X							X	LOW

BA, Bland-Altman analysis; CCC, concordance correlation coefficients; ECG, electrocardiography; ICC, interclass correlation; MAE, mean absolute error; MAPE, mean absolute percent error; ME, mean error; MPE/RER, mean percent error or relative error rate; PC, Pearson correlation; VO_{2max}, maximal oxygen uptake.

Table 2.

Summary of study findings

Study author/year	Sample population	Design/methods	Summary of findings
Hwang J, et at. (2019)	N=51, mean age 44.4 (SD 16.6) yr, n=24 female, Korean	Compared FBC2 to ECG; participants with history of paroxysmal supraventricular tachyarrhythmia (SVT) were undergoing electrophysiological study; baseline HR during induced SVT and post ablation of SVT were measured	FBC2 accurately measures HR (94%) compared to ECG
Klepin K, et at. (2019)	N=60, mean age 31 (SD 7.3) yr, n=33 female, race not reported	Compared FBC2 to VO_{2max}; all healthy adults; study duration: 1-week; 3 GPS tracked 15-minute outdoor runs and were worn continuously	FBC2 60-second cardiorespiratory fitness (CRF): highly associated with VO_{2max}; MAPE 9.14%
Nelson BW, et al (2019)	N=1, age 29 yr, white male	Compared FBC2 to ECG; single subject, healthy adult male, used to minimize variable differences; 24-hour data collected across sedentary, walking, running, activities of daily living (ADLs), and sleeping	FBC2 mean difference −3.47 bpm compared to criterion and comparable tracker; MAE 5.96%, CCC 91% compared to ECG over 24 hours; MAPE 9.21% walking
O'Driscoll R, et al. (2019)	N=59, mean age 44.2 (SD 14.1) yr, n=41 female, European	Compared FBC2 to Polar chest strap; activities evaluated in a 1-day laboratory setting included running, walking, cycling, mimicked ADLs, and sedentary conditions with rest periods between	FBC2 HR = CM, but not consistent across activity levels; MAPE 31% walking incline; MAPE 69% walking
Reddy RK, et al. (2018)	N=20, mean age 27.5 (SD 6.0) yr, n=11 females, n=17 (85%) white	Compared FBC2 to Polar chest strap; 2-day data collection; all healthy adults; VO_{2max} testing included resistance exercise, interval training, and ADLs conditions	FBC2 >error high intensity exercise; MAPE 10.79%
Tedesco S, et al (2019)	N=20, mean age 70.2 (SD 2.9) yr, n=11 female, all white (English/Irish)	Compared FBC2 to ActiGraph; 24-hour free-living data collection; all older, healthy adults; activities assessed included a range of moderate to vigorous walking and sleeping conditions	All devices highly correlated (ICC>0.89), FBC2 overcounted steps, MAPE 12.36%
Thomson EA, et al (2019)	N=30, mean age 23.5 (SD 3) yr, n=15 female, race not reported	Compared FBC2 to ECG; measurements taken at intervals: rest (3 min), standing (2 min), treadmill (every 3 min with gradual speed and uphill increase until volitional fatigue), recovery.	FBC2 error rate (3.9–13.5%). RER per activity level: light (5.36%), moderate (9.20%), vigorous (11%).
Vetrovsky T, et al. (2019)	Healthy control participants: n=15, mean age 65.5 (SD 12.6) yr, n=6 female. HF field-based study: n=14, mean age 43.3 (SD 18.9) yr, n=9 female. (Czech Republic)	Compared FBC2 to accelerometer; main purpose was to evaluate step accuracy of activity trackers in persons with HF; laboratory and field study	FBC2 healthy participants MAPE 12%; HF study MAPE 46%; >correlation low speeds on treadmill